Systemic Bias of IMF Reserve and Debt Forecasts for Program Countries

Theo S. Eicher
University of Washington

and

Reina Kawai
University of Washington

Countries experiencing balance of payments (BOP) crises may obtain IMF loans to stabilize external accounts. These loans require IMF programs that outline performance targets to ensure forecasted recovery trajectories. Two key indicators of external account performance are reserves and short-term external debt ("STdebt"). An extensive literature evaluates IMF forecasts, but reserves and STdebt have not been studied to date. We construct a database of nearly 300 BOP crisis countries with IMF BOP programs from 1992-2019. Reserve forecasts are shown to be systematically biased and inefficient, a result that is startlingly persistent across (a) degrees of capital mobility, (b) trade openness, (c) exchange rate regimes, (d) inflation, and (e) country income levels. We show the bias is driven by deeply pessimistic IMF reserve forecasts that underestimate reserves and systematically ignore information known at the time of the forecast. STdebt forecasts are also inefficient but with an optimistic bias, systematically underestimating future debt. If STdebt is used to peg reserve requirements, the optimistic bias of STdebt forecasts may drive the pessimistic bias of reserve forecasts.

Keywords: *IMF Forecasts, BOP Crises, Reserves, Short-Term External Debt*

JEL Classifications: *O19, O11, F47*

We thank the IMF for providing access to the Monitoring of Fund Arrangement (MONA) database and archived Executive Board Loan Program documents. Chris Papageorgiou, Charis Christofides, Monica Gao Rollinson, Monique Newiak, and David Kuenzel provided helpful discussions in a previous, related project. Chengjun Zhang, Liyuan Zhang, Elena Zhu, Ziqing Wang, and Xuanming Da provided research assistance. An anonymous referee provided valuable suggestions.

I) Introduction

A voluminous literature studies the adequacy of countries' reserve levels to avert balance of payment (BOP) crises.[1] This paper takes a different tack that has not been investigated to date. We examine whether countries experiencing BOP crises can rely on unbiased and efficient IMF forecasts for reserves and short-term external debt to chart their economic recovery trajectories.

During BOP crises, insufficient reserves and excessive external STdebt undermine exchange rate regimes, debt sustainability, and economic stability. Countries may obtain financial assistance loans from the IMF to stabilize external accounts. IMF loan disbursements are, however, tied to detailed performance criteria ("conditionality") laid out in a program that charts a recovery trajectory. Along this trajectory, Reserve Coverage (gross official reserves in months of imports) has long been a key indicator of program performance.[2] Bordo et al (2004), for example, rank Reserve Coverage as one of four canonical macroeconomic performance indicators in IMF loan programs. Following the Asian Financial Crisis, the ratio of short-term external debt to reserves was added as an additional key indicator to peg proper reserve levels for BOP crisis countries.[3]

Historically there has been great interest in evaluating the accuracy of forecasts for key IMF program indicators.[4] Surprisingly, however, the accuracy of IMF reserve or debt forecasts has not been scrutinized. To do so, we audit the IMF's *Monitoring of Adjustment Lending Database* (MONA), correct errors, and fill in missing data from IMF archival loan documents (the original source of MONA data). This allows us to construct a novel and unique database of 287 IMF BOP programs over 29 years, from 1992-2020. The only previous study that examined IMF reserve forecast accuracy for program countries is Musso and Phillips (2002), who covered 65 programs over 4 years (1993-1997).[5] The size of our dataset increases the power of our inference and allows us to probe deeper than Musso and Phillips (2002). The

[1] See Aizenman and Genberg (2012) for an excellent survey.
[2] Reserve Coverage is also interchangeably referred to as "Import Cover" or "Import Coverage." These terms possess analogs with different meanings in the tariff literature, hence we use Reserve Coverage.
[3] Recently, IMF programs also report at times broader reserve adequacy statistics which we discuss below, however, Reserve Coverage and short-term debt remain central, especially for (i) low-income countries, (ii) countries with limited capital mobility, (iii) countries with low trade shares. See IMF (2011), IMF (2013), IMF (2015), IMF (2016), and Ferrer and Kireyev (2022).
[4] A voluminous literature, spanning several dozen publications, examines potential bias and inefficiency of IMF forecasts for GDP, inflation, current account, and fiscal indicators in program countries. See Genberg and Martinez (2014) and Eicher and Gao Rollinson (2022) for surveys.
[5] IMF (2019) also presents an evaluation of reserve forecast accuracy of sorts, by visually inspecting a forecast error histogram, to conclude that reserves "IMF-supported programs ongoing during the period September 2011 to end-2017… met or came close to reaching projected levels."

dataset is of sufficient size to examine whether bias and inefficiency differs by exchange rate regimes, levels of development, inflation, capital account openness, trade openness, or across time.

Previous studies evaluating IMF forecasts established a systematic optimistic bias for GDP, inflation, fiscal and external balances, always underestimating eventual outcomes.[6] Instead, we document that Reserve Coverage forecasts exhibit systematic *pessimistic* bias and inefficiency. The result is shown to be startlingly persistent across degrees of capital mobility, trade openness, exchange rate regimes, income levels, and inflation levels. We untangle Reserve Coverage forecast inaccuracies by imports and reserves, to find the bias and inefficiency is driven by deeply pessimistic reserve forecasts that systematically ignore information known at the time of the forecast. Importantly, we show that the economic effects of IMF conditionality are not properly integrated into IMF forecasts, although conditionality is known at the time of the forecast. Effects of conditionality that are improperly accounted for in forecasts are BOP tests, limits on government credit/deficit, and limits on arrears. In addition, we show that noneconomic events, known at the time of the forecast, are also not properly integrated into IMF forecasts. Specifically, disasters, international conflicts, and civil wars. These systematic errors should be addressed with additional caution in future IMF forecasts for program countries to improve the accuracy of projected recovery trajectories.[7]

The canonical metric for BOP sustainability, Reserve Coverage, focuses on "internal drains." Following the Asian Crisis, the IMF expanded the metric to include BOP sustainability indicators, especially short-term external debt ("STdebt") as a proxy for liquidity risk and capital account vulnerabilities (IMF 2000). Bordo et al (2004) suggest a mix of Reserve Coverage and Debt Coverage (reserves-to-STdebt) could be used for program countries.[8] To understand the underlying drivers of reserve forecasts, we also examine IMF STdebt forecasts below. Since STdebt is infrequently reported for programs prior to 2002, our debt dataset consists of 124 program observations from 1992-2019. To our knowledge, it is the first dataset and evaluation of the IMF's STdebt forecasts for BOP crisis countries.

[6] See Genberg and Martinez (2014), Carrière-Swallow and Marzluf (2021), Ismail, Perrelli, and Yang (2020), Beaudry and Willems (2018), Celasun, Lee, Mrkaic, and Timmermann (2021), Artis (1988 and 1997), Barrionuevo (1993), Timmermann (2007), Baquir, Ramchran, and Sahay (2005), Dreher, Marchesi, and Vreeland (2007), Baker and Rosnick (2003), Eicher and Gao (2022), Eicher and Kawai (2022), Eicher, Kuenzel, Papageorgiou, and Christofides (2019).

[7] We also examine whether our results are driven by influential or extreme values and find results are robust.

[8] Post 2008, the IMF designed a new composite reserve indicator to be "sensitive to the stylized economic structures" of individual countries. In addition to short-term external debt, the IMF ARA metric considers "other medium/long term debt liabilities, broad money, and exports" (IMF 2016). The metric is at times reported in loan programs after 2014, but its exact composition and underlying data for each particular country is seldom provided.

Similar to our findings for Reserve Coverage, STdebt forecasts in IMF programs are shown to be systematically inefficient and usually biased across country income levels, exchange rate regimes, and financial/trade openness. Curiously, however, STdebt is systematically *optimistically* forecasted (except for fixed exchange rate regimes). This implies programs systematically suggest BOP crisis countries will accumulate much lower debt along their recovery path than what is observed in the actual, final data. The fact that bias and inefficiency are highly statistically significant and systematic indicates that we are not documenting random data errors or effects of idiosyncratic events; such errors would have averaged out over time, especially in a dataset of the size and duration as ours. It is important to note that falsely optimistic, excessively low-forecasted debt levels may drive some of the false Reserve Coverage pessimism that we observe. STdebt forecasts with systematic optimistic bias imply lower than necessary reserves to cover debt in case of sudden stops in capital flows.

We undertake a special effort to identify factors that are known at the time of the forecast but which may not have been properly integrated into forecasts. Similar to Musso and Phillips (2002) we find that loan size is a key factor that biases IMF reserve forecasts. Musso and Phillips use a general-to-specific model selection criterion to establish their result (see Fernandez et. al., 2001 and Raftery, 1995). Instead we use Sinclair et al.'s (2010) extension of Mincer and Zarnowitz's (1969) statistical approach. We also examine a wider range of factors such as a) IMF program conditionality, b) program size, c) crises (Global and Asian), and d) exogenous/endogenous events (elections and conflict).[9] Results indicate that reserve bias is driven by the systematic, improper integration of the IMF's own conditionality involving (a) BOP tests, (b) limits on government credit, (c) program size, as well as the influence of noneconomic events such as disasters and conflicts.

Below we proceed as follows: Section II presents data and methodology. Section III examines reserve coverage bias and inefficiency by subsamples (country incomes, exchange rate regimes, inflation, capital mobility, and trade shares) and decomposes forecast errors into reserve level and import forecasts. Section IV investigates if Reserve Coverage forecast accuracy has changed over time. Section V explores sources of bias and inefficiency of IMF STdebt forecasts by subsamples and across time. Section VI examines drivers of reserve and debt forecast inefficiencies. Section VII explores robustness, and section VIII concludes.

[9] Economic events are "endogenous" when they took place within a year before the start of a program. Exogenous events occur after the program started

II) Data and Methodology

II.1) Data

IMF forecasts for countries with BOP programs were obtained from the IMF's *Monitoring of Fund Arrangements* database (MONA, IMF, 2021a).[10] The MONA database reports economic indicators from loan documents that the IMF's Executive Board approved for each crisis country.[11] We audit the MONA database, correct 110 errors and fill in 87 missing reserve coverage data points using IMF's archival loan documents; a list of errors/corrections is provided in Appendix C. Our focus is on one-step-ahead crisis-year forecasts (forecasts in crisis year *t*, for year *t)*, which implies that we are examining only the most immediate forecast that where formed using the most recent program design data. Missing observations in the MONA database and in the loan documents make the assessment of forecast accuracy for longer forecast horizons more tentative, we explore them in the robustness section VII below.

We examine Reserve Coverage, defined as the level of gross international reserves (in USD million) over the monthly level of imports of goods and services (total imports/12 in USD million).[12] We also examine STdebt (in USD billion), defined as external debt with maturities of less than 12 months. STdebt is more frequently held confidential in loan documents than reserves, hence the debt sample is slightly smaller. At times IMF documents reference the ratio of reserves to STdebt as a proxy for capital market exposure. Instead below we focus on forecasts of STdebt levels since many countries report zero STdebt, which would render the ratio infinite. In recent years IMF programs focus predominantly on net reserves – these are, however, not consistently calculated and/or reported in program documents, especially prior to 2002. Hence we focus on gross international reserves.

Data on STdebt is well known to be subject to significant margins of error (IMF 2000). When unavailable in MONA, we collect 43 STdebt observations from IMF archival loan

---

[10] IMF loan programs that focus on BOP crises include Extended Credit Facility (ECF), Extended Fund Facility (EFF), Exogenous Shock Facility (ESF), Flexible Credit Line (FCL), Stand-By Agreements (SBA), Standby Credit Facility (SCF), Precautionary Credit Line (PCL), Precautionary Liquidity Line (PLL), as well as ECF-EFF, SAB-SCF, and SBA-ESF programs.

[11] The IMF's World Economic Outlook (WEO) database provides forecasts in April and October of each year, but it does not provide forecasts for countries in crisis. WEO forecasts for individual developing countries are not available prior to 2004.

[12] Alternatively, we could conduct the analysis in *growth rates* or *log changes* (*t-1 to t*). Results are qualitatively similar and discussed in robustness section VII.4. We prefer Reserve Coverage as our unit of analysis which is (a) the data reported in IMF loan documents, and (b) the basis for the canonical "Reserves Coverage" rule of thumb that is prevalent in IMF lending and often referred to in loan documents. Net International Reserves are another alternative, but these do not relate to the popular "Reserves Coverage" metric, and lack a standard definition in IMF Balance of Payments Manuals.

documents for these programs. Final outcome STdebt data was obtained from the World Bank because IMF STdebt data is confidential. We had access to IMF STdebt data and found World Bank and IMF debt data are highly correlated (correlation coefficient of 0.89). We lose 5 observations because we are unable to include confidential IMF data, but the inclusion of confidential IMF STdebt data generates very similar results as we report below.

As we prepared this study, the IMF's *MONA Database* reported data for 324 BOP crisis programs in 113 countries over 29 years, from 1992 to 2020. After auditing the database and IMF archival documents, we managed to produce a database with 287 Reserve Coverage observations and 124 STdebt observations. Our dataset is thus over 4 times larger and 25 years longer than the sole previous IMF reserve forecasts evaluation for program countries (Musso and Phillips, 2002). To our knowledge, an evaluation of forecasts for STdebt in program countries does not exist to date.

The size of our dataset allows us to examine the accuracy of IMF forecasts across a number of important subsamples. We examine fixed and flexible exchange rate regimes, where our "fixed"/"flexible" definitions follow Ilzetzki et al. (2019) and the exchange rate regime classifications of the IMF's *Annual Report on Exchange Arrangements and Exchange Restrictions* (IMF 2021e). Ilzetzki et al's (2019) coarse exchange rate regime classification, identifies fixed exchange rates as i) no separate legal tender, ii) preannounced pegs, iii) currency boards, iv) de facto pegs, and v) preannounced horizontal bands ≤ 2%. We also examine forecast accuracy by capital account openness and trade shares. For capital account openness we use the Chinn-Ito index, which has by far the largest coverage for our sample, and classify countries as "high capital mobility" when a country's index exceeds the mean. The use of the Chinn-Ito index reduces our sample by 31 observations. For trade openness, we calculate GDP trade shares using the World Bank's (2022a) *National Accounts Database* and classify countries' trade openness as "high" when a country's trade share exceeds the mean. Finally, we separate the sample by income levels and classify countries according to the World Bank's (2022b) time-variant *Country Lending Group Income Classification*. To identify hyperinflation countries, we follow Dornbusch and Fischer's (1986) 25% inflation threshold.

Actual final data was obtained from three official databases. Final Reserve Coverage data was obtained from the World Bank's (2022d) *World Development Indicator Database* which cites the "International Monetary Fund, International Financial Statistics and Data Files" as its sources. When reserve data was missing, we augmented WB data with reserve data from the IMF's *Assessing Reserve Adequacy Database* (IMF, 2022a) for 13 programs. Final imports of goods and services were obtained from the IMF's *International Financial Statistics Database* (IMF, 2021c), and final STdebt data was obtained from the World Bank's *World Development Indicator Database* (WB, 2022d) which cites "World Bank, and the World Bank's *International Debt Statistics Database*" (WB, 2022c) as its source.

We also examine whether forecasts systematically ignore information known at the time of the forecasts. Here we introduce data on elections, conflicts, crises (Global and Asian) and on IMF conditionality. We distinguish between "endogenous" events that occurred before the forecasts were established (e.g., "endogenous wars") vs "exogenous" events that occurred *after* the start of the program and hence unknown to forecasters (e.g., "exogenous disasters"). For disasters, we use the OFDA/CRED's *International Disaster Database* (EM-DAT, 2000). For election data, we use Beck et al.'s (2001) data pre1998 and IFES's (2020) post-1998 data on head-of-state and legislative elections. For conflict data, we use Harbom et al.'s (2009) data on intra/inter country conflicts. For conditionality, we use the IMF's *MONA Database*, which includes the IMF's own coding of conditionality. The MONA Glossary (IMF 2021a) explains that the IMF groups conditionality into 11 different subcategories: Domestic Credit Ceiling, Gov't/Public Sector Credit Ceilings, BOP Reserve Tests, Debt Ceilings (short, medium and long term), Arrears Ceilings (domestic and external), Fiscal Deficit Ceilings. Appendix B provides a description of our data.

II.2) Evaluating Forecast Accuracy: Methodology

A frequently used metric to evaluate forecasts is the Mean Absolute Error (MAE), or variants thereof that adjust for percentages (e.g., Mean Absolute Percentage Error), black swan events with unusually large errors (e.g., Root Means Square Error), or scale invariance (e.g., Mean Absolute Scaled Error). Such metrics are useful when studies compare the *relative* forecast accuracies of different forecasters (e.g., OECD vs IMF vs WB forecasts). The nature of BOP crises and IMF crisis programs dictates that our paper can only examine the accuracy of a single forecast. Only the IMF has access to country-level data in times of crisis, hence it is the only entity that can establish a forecast in times of crisis. This highlights the importance of the accuracy of IMF forecasts: they are required for the lender of last resort's loan approval documents, and they will be the benchmark for future loan disbursements. The fact that only a single forecast source is available for forecasts in times of crisis presents some challenges for forecast evaluation since we cannot utilize *relative* forecast accuracy to inform forecast quality.

In our application, an assessment of forecast accuracy requires formal statistical tests. These tests were first developed by Mincer and Zarnowitz (1969), who extended the seminal work of Theil (1961). Theil developed a "*Prediction-Realization Diagram*" with forecasts for the current year, F_t, on the horizontal axis and official, actual final data for the current year, A_t, on the vertical axis. The 45-degree line of the Prediction-Realization Diagram is the "*Line of Perfect Forecasts*," representing coordinates where forecasts coincide with actual final data. The *Mean Square Forecast Error*, $E(A - F)^2$, is simply the variance around the Line of Perfect Forecasts, which motivates standard regression techniques to evaluate forecast accuracy.

Formal tests for unbiased and efficient forecasts were suggested by Mincer and Zarnowitz (1969) by estimating

$$A_t = \alpha + \beta F_t + \varepsilon_t, \tag{1}$$

where forecasts are chosen as the "independent" variable in (1) only because forecasts are available before the actual final data is published. Forecasts are thought to be efficient when forecast errors are random and uncorrelated with forecasts. The concept of efficiency here is akin to Nordhaus's (1987) stock market efficiency – both cases require that all available information is reflected in the forecast, and that errors are white noise. Mincer Zarnowitz suggest forecast efficiency requires a test with the joint null hypotheses, $\alpha = 0$ and $\beta = 1$.

When the Mincer-Zarnowitz null hypothesis of $\alpha = 0$ & $\beta =1$ is rejected, forecasts are inefficient due to excessive deviations from the line of perfect forecast. Inefficient forecasts may, however, not necessarily be biased. The concept of bias represents a lower accuracy threshold as it requires only that averages of forecasts and final data are not significantly different, $E(A_t) \neq E(F_t)$. Holden and Sandhu (1987) demonstrate that $\alpha = 0$ is sufficient, but not necessary for unbiased forecasts since forecasts can be "unbiased" even when, for example, half are 40 percent higher, and half are 40 percent lower than the actual final data. Holden and Peel (1990) derived necessary and sufficient conditions for bias, examining whether the regression line intersects the *Line of Perfect Forecasts* at $E(A_t) = E(F_t)$. When the regression $A_t - F_t = \gamma + v_t$ rejects the null of $\gamma = 0$, forecasts are said to be biased.

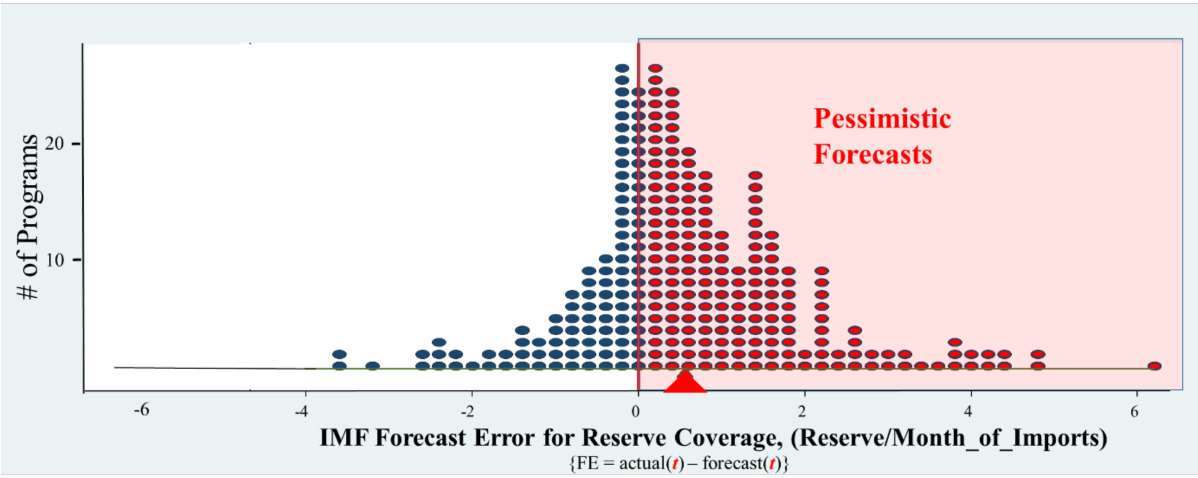
The Mincer Zarnowitz approach to forecast evaluation has a long tradition in IMF forecast accuracy evaluation. It was first employed by Kenan and Schwartz (1986) to evaluate IMF forecast accuracy for WEO variables. The technique has since then been applied frequently in IMF forecast evaluations by Artis (1996), Musso and Phillips (2002), Timmermann (2007), Genberg and Martinez (2014), Eicher et al (2019), and Eicher and Gao Rollinson (2022) and Eicher and Kawai (2022).

III) Bias and Inefficiency of IMF Reserve Coverage Forecasts

We commence with an analysis of Reserve Coverage forecast accuracy and address STdebt forecasts in Section V below. Figure 1 displays a histogram of Reserve Coverage forecast errors, $A_t - F_t$, for our 287 BOP program countries from 1992-2020. The figure indicates that, on balance, Reserve Coverage forecasts are pessimistic. Pessimism here expresses that the IMF programs forecast lower Reserve Coverage than actual final data reports.

While Figure 1 indicates that forecasts are on balance pessimistic, it remains unclear whether the bias is actually statistically significant and/or perhaps inefficient. The only evaluation of IMF Reserve forecasts aside from Musso and Phillips (2002) bases its assessment on a histogram akin to Figure 1. The authors of the IMF's *Review of Program Design and Conditionality* (IMF 2019) use visual inspection of a histogram to assert that "reserve targets were met or came close to reaching projected levels."

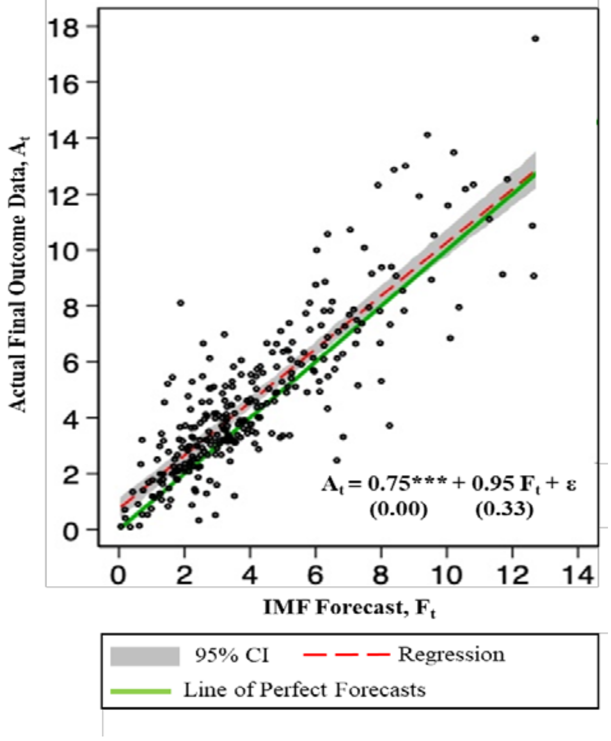
Figure 1
Reserve Coverage Forecasts
287 IMF BOP Crisis Programs, 1992-2020



Using the Mincer Zarnowitz statistical approach to forecast evaluation, the Prediction-Realization diagram for Reserve Coverage forecasts is provided in Figure 2 along with the line of perfect forecasts (green). The figure also includes the Mincer Zarnowitz regression line (1) in red, which minimizes the distance between predicted and actual values. The associated 95% confidence interval is drawn in gray. The p-value of the forecast coefficient, β, is associated with the tests of $\beta = 1$ (not the standard $\beta = 0$), following the Mincer Zarnowitz null hypothesis that forecasts are unbiased and efficient.

Several important observations are of note in Figure 2. First, despite the fact that the coefficient estimate on IMF forecasts is less than unity, $\beta = 0.95$, the regression line never intersects the line of perfect forecasts over the entire range of forecasts, indicating a lack of efficiency. This is due to our second observation that the intercept is positive and substantial, indicating systematic, pessimistic bias of Reserve Coverage forecasts. Third, the dispersion around the line of perfect forecasts is substantial reflecting a lack of forecast accuracy. We examine the formal statistical properties of IMF Reserve Coverage forecasts in the next section. Third, the intercept of 0.75 indicates that countries with the greatest BOP crises (whose Reserve Coverage had contracted to the lowest levels) are also the countries that receive the most pessimistic IMF Reserve Coverage forecasts. To quantify this disparity, we examine "high BOP crisis countries" (with Reserve Coverage forecasts below the 4.1 Reserves/months of imports sample mean) to find their Reserve Coverage is underestimated by an average of 25%. For "highest BOP crisis countries" (Reserve Coverage forecasts at the lowest quartile, 2.3 months of reserves/months imports) forecasts are on average, 55% lower than the actual observed data.

Figure 2
Prediction-Realization Diagram Forecasts vs Actual Outcomes
Reserve Coverage (Reserves/Months of Imports)



III.1) Baseline Results

Results of the Mincer Zarnowitz regressions in equation (1) for IMF Reserve Coverage forecasts across different country samples, along with formal statistical tests for bias and inefficiency are reported in Table 1. We report full sample results along with sub-samples by (i) levels of development (All, LICs, Non-LICs), (ii) inflation, and (iii) exchange rate regimes. Different levels of development are associated with different levels of global financial and commercial integration. High-inflation countries are notoriously difficult to forecast and these forecast errors may contaminate the results of the full sample. Finally, reserves serve profoundly different purposes in fixed and flexible exchange rate regimes, hence one might suspect that forecast accuracy may differ across these subsamples.

Overall, the results indicate remarkably biased and inefficient Reserve Coverage forecasts across all samples. For the full sample, all country-income subsamples, all inflation subsamples, and all exchange rate subsamples, Reserve Coverage forecasts are statistically significantly biased (Holden Peel test (HP), p-values provided) and inefficient (Mincer Zarnowitz test (MZ), p-values provided). The null hypothesis of unbiased and efficient forecasts is rejected at significance levels that exceed 0.01, with the exception of fixed exchange rate regimes where significance levels exceed 0.05 levels. For LICs with fixed exchange rates (a subsample hampered by its concerning small size, 16 observations), the Mincer and Zarnowitz and Holden Peel tests produce unbiased and efficient forecasts. We would not want to over-interpret this finding given the small sample size. Table 1 also features large positive and highly statistically significant intercepts that exceed actual reserves by about 1 month throughout.[13]

[13] An alternative definition for "LICs" countries could be the IMF's LICs definition which relies on country access to concessional financing (poverty reduction and growth trust, PRGT, see IMF 2020). While the World Bank classification relies on income, the IMF classification focuses also on the ability to borrow from international financial markets on a durable and substantial basis. The IMF definition produces a larger set of LICs but identical results with the exception of fixed exchange rate LICs. That sample is now large enough to provide sufficient power to also reject unbiased and efficient forecasts at the 1 percent level.

Table 1
Bias and Inefficiency of IMF Reserve Coverage Forecasts
By Income Levels, Exchange Rate Regimes, and (Hyper) Inflation

Reserve Coverage	**ALL**			**Non-Hyper**			**LIC**			**Non-LIC**		
Dependent: Final Data	**All**	**All Float**	**All Fixed**	**All**	**All Float**	**All Fixed**	**All**	**All Float**	**All Fixed**	**All**	**All Float**	**All Fixed**
Forecast (β)	0.95	0.97	0.92	0.96	1.00	0.91	1.05	1.21	0.94	0.93	0.94	0.91
p-value (β=1)	0.33	0.65	0.21	0.50	0.95	0.22	0.62	0.13	0.60	0.24	0.41	0.28
Constant (α)	0.75***	0.76***	0.73***	0.71***	0.64**	0.82**	0.47	0.07	0.68	0.82***	0.87***	0.76**
p-value (α=0)	0.00	0.00	0.01	0.00	0.01	0.02	0.12	0.83	0.11	0.00	0.00	0.04
Observations	287	191	96	241	156	85	58	42	16	229	149	80
Adj. R^2	0.735	0.743	0.720	0.740	0.765	0.693	0.819	0.788	0.900	0.712	0.732	0.661
MZ p-value (α=0 & β=1)	0.00***	0.00***	0.01**	0.00***	0.00***	0.01**	0.00***	0.00***	0.13	0.00***	0.00***	0.04**
HP p-value (γ=0)	0.00***	0.00***	0.01**	0.00***	0.00***	0.01**	0.00***	0.00***	0.12	0.00***	0.00***	0.04**

Robust p-values reported, *** p<0.01, ** p<0.05, * p<0.1

Table 1 highlights that results are not driven by hyperinflation countries, as coefficients and significance levels are similar in the Non-Hyperinflation sample; this eliminates concerns about possible contamination from difficult to forecast hyperinflation programs. Instead, bias and inefficiency are squarely driven by Non-LICs, a subsample that is overwhelmingly dominated by emerging market MICs (Non-LICs include only 7 HICs programs). Surprising is that Reserve Coverage is equally inaccurately forecast for fixed exchange rate regimes in the full, Non-LICs, and Non-Hyperinflation samples. Floating and fixed exchange rate regimes face somewhat different challenges during BOP crises, nevertheless we do not observe differential forecast accuracy for Reserve Coverage.

III.2) Possible Interpretations of Baseline Results

It is tempting to suggest low reserves and little room to maneuver during BOP crises may induce the IMF to design reserve buffers into programs by issuing systematically pessimist forecasts. This hypothesis could also explain that, the worse the crisis, the greater the caution and hence the larger the buffer (bias) that is observed. On the other hand, it is puzzling that deliberate, systematic risk buffers which increase with the severity of crises, are not mentioned by the IMF reserve adequacy guidance literature or in loan documents.

Another explanation may be the presence of an anchoring effect, a systematic cognitive bias whereby forecasters are influenced by a particular reference point or “anchor.” The anchor in this case would be net international reserves, which are at times subject to IMF conditionality. If net international reserves are specified as part of “reserve tests” in IMF loan documents, they involve a particular “net reserve floor” that cannot be pierced as a loan memorandum item. If gross reserves forecasts are then projected off the net international reserve floor, forecasters may be systematically anchoring actual gross reserves too low. We investigate this possibility further in Table 6 below, when we examine which type of IMF conditionality might contribute to systematic bias. Indeed, Table 6 presents statistical evidence that both Reserve Coverage and Reserve Levels exhibit more systematic, pessimistic bias when reserve test conditionality in terms of net reserve floors are part of a program.

Alternatively, Musso and Phillips (2002) suggest pessimistic IMF reserve forecasts may relate to the structure of the IMF Executive Board loan review process. They worry the review process may incentivize staff to forecast initially pessimistic outcomes to characterize future results as “unexpectedly better” and to avoid “unexpectedly weak” outcomes. This explanation does not address why pessimism changes with reserve levels.

Most important are, however, the real implications of our finding of systematic bias and inefficiency. There are costs to pessimistic forecasts and program design, as well as to accidental and/or implicit/planned buffers. First, one would have to question why this highly statistically significant pessimism and risk buffer is absent for the economically most vulnerable countries: LICs. Table 1 reveals that for LICs the intercept is not statistically different from zero. Secondly, there are real costs for countries that exceed their required reserve accumulation. A sizable literature discusses the opportunity costs of accumulating greater than specified Reserve Coverage affects the efficacy of the program design. Funds used for reserve accumulation could be used to reduce the impact on domestic absorption. Or these funds could provide crisis remediation in other sectors of the economy.

III.3) Capital Account and Trade Openness as Drivers of Bias and Inefficiency

Throughout the 1990s, BOP crisis metrics for IMF program countries focused on “internal drains” captured by Reserve Coverage (Wijnholds and Kapteyn, 2001). Even today, Reserve Coverage is reported in just about all IMF programs, and the rule of thumb that reserves should cover at least three months of imports is still firmly ingrained (see e.g., IMF 2022b and Chitu, et. al., 2019). Indeed, this rule of thumb is still applied informally, even to advanced countries with high degrees of capital mobility, for example during the 2022 British fiscal crisis, see Bloomberg (2022). After the Asian Financial Crisis, the IMF developed more granular reserve metrics, depending on countries' "*domestic financial development; credibility of monetary policy, and the exchange rate regime; whether the economy is (unofficially) dollarized or euroized; its trade and financial openness*" (see Chitu, et. al., 2019). Aizenman and Genberg (2012) survey the literature and provide an overview of the theoretical and empirical considerations to vary reserve levels by openness. Fundamentally, capital openness allows for greater market access and therefore diminished need to fully self-finance reserve needs. On the other hand, capital openness also provides for hot money and greater pass-through of global fluctuations to shock the domestic economy. It is thus easily conceivable that different levels of openness provide different forecasting challenges. Having searched for possible differences in forecast accuracies by levels of development, exchange rate regimes, and inflation, we thus broaden our focus to include the impact of openness (trade and financial) on forecast accuracy.

Table 2 reports that systematic bias and inefficiency are invariant to either trade or financial openness. Forecasts remain pessimistic throughout, with intercept estimates exceeding a month for imports at times. For low capital mobility countries, the coefficient on forecasts is now below unity and statistically significant, and it combines with large and significantly positive intercepts. This indicates not only the invariance of forecast inaccuracy to openness but also that pessimistic underestimates of Reserve Coverage are especially pronounced in low capital mobility countries and in countries that face the greatest BOP crises.

Table 2
Bias and Inefficiency of IMF Reserve Coverage Forecasts
By Trade and Capital Account Openness

Reserve Cover	**KA High**				**KA Low**				**Trade High**				**Trade Low**			
Dependent: Final Data	**ALL**	**Non-Hyper**	**Non-LIC**	**LIC**	**ALL**	**Non-Hyper**	**Non-LIC**	**LIC**	**ALL**	**Non-Hyper**	**Non-LIC**	**LIC**	**ALL**	**Non-Hyper**	**Non-LIC**	**LIC**
Forecast (β)	0.97	0.95	0.98	1.05	0.79***	0.81***	0.72***	0.93	0.91	0.87	0.89	1.13	0.95	0.98	0.92	1.04
p-value ($\beta=1$)	0.71	0.51	0.79	0.85	0.00	0.00	0.00	0.43	0.21	0.10	0.12	0.43	0.39	0.74	0.30	0.73
Constant (α)	0.82***	0.85***	0.68**	1.12	1.14***	1.10***	1.47***	0.48**	0.74***	0.93***	0.81***	0.19	0.85***	0.70***	0.97***	0.53
p-value ($\alpha=0$)	0.01	0.01	0.04	0.28	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.61	0.00	0.01	0.00	0.14
Observations	102	95	88	14	154	126	114	40	110	92	96	14	177	149	133	44
Adj. R^2	0.75	0.76	0.77	0.64	0.69	0.70	0.61	0.88	0.68	0.65	0.66	0.87	0.73	0.75	0.69	0.81
MZ p-value ($\alpha=0$ & $\beta=1$)	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.02**	0.00***	0.00***	0.00***	0.03**	0.00***	0.00***	0.00***	0.00***
HP p-value ($\gamma=0$)	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.01**	0.06**	0.00***	0.00***	0.00***	0.02**	0.00***	0.00***	0.00***	0.00***

Robust p-values reported, *** p<0.01, ** p<0.05, * p<0.1

Tables 1 and 2 provide a surprisingly uniform and robust picture that Reserve Coverage for IMF BOP program countries is forecasted with significant bias and inefficiency, no matter the level of development, exchange rate regime, inflation, trade openness or financial account openness. The next step is thus an effort to dig deeper and decompose the Reserve Coverage ratio into reserve forecast errors and import forecast errors to illuminate whether Reserve Coverage forecast inaccuracies are driven by either reserves and/or imports. Below we examine the accuracy of reserve and import forecasts in levels to match Reserve Coverage data in IMF loan documents.

III.4) Composite Indicators Produce Composite Forecast Bias and Inefficiency

Since the Reserve Coverage indicator is a ratio, its forecast effectively represents a composite forecast of reserve levels and import levels. Given the secular bias and inefficiency of Reserve Coverage forecasts across subsamples, it is natural to ask if bias and/or inefficiency is driven by inaccuracies in reserve and/or import forecasts. The results for the two individual forecasts are provided in Table 3. Two important insights emerge. First, bias and inefficiency emanate overwhelmingly from forecasts of reserve levels, not from import forecasts. This result mirrors the finding of Eicher and Kawai (2022) who document that import growth is unbiased and efficiently forecast for IMF program countries. Second, Table 3 starkly emphasizes the problems associated with the use of composite indicators: the sub-sample where reserve forecasts are not biased and inefficient (fixed exchange rate regimes) is exactly the (only) sample that is inefficient for import forecasts. Although the reserve and import forecasts are not both inefficient across subsamples, their individual inefficiencies combine to produce the biased and inefficient forecasts in *all* sub-samples of Reserve Coverage observed in Table 1.

Only the subsample for LICs with fixed exchange rate regimes is unbiased and efficient across reserve *and* import forecasts, to produce unbiased and efficient Reserve Coverage forecasts for that subsample. As in the case of Reserve Coverage, we view this LICs result with some caution, given the exceedingly small sample size.

Table 3a
Bias and Inefficiency of IMF Forecasts
Reserve Levels and Import Levels

Reserve Levels	**Full Sample**			**Non-Hyper**			**LIC**			**Non-LIC**		
Dependent: Final Data	**All**	**Float**	**Fixed**	**All**	**Float**	**Fixed**	**All**	**Float**	**Fixed**	**All**	**Float**	**Fixed**
Forecast (β)	0.99	0.98	1.00	0.98	0.98	1.00	1.01	1.02	0.93	0.98	0.98	1.00
p-value (β=1)	0.49	0.45	0.97	0.34	0.31	0.98	0.89	0.75	0.45	0.45	0.39	0.94
Constant (α)	565***	765***	115	576***	772***	131	152**	169*	161*	671***	927***	124
p-value (α=0)	0.00	0.00	0.37	0.00	0.00	0.39	0.06	0.10	0.09	0.00	0.00	0.42
Observations	287	191	96	241	156	85	58	42	16	229	149	80
Adj. R^2	0.97	0.97	0.96	0.97	0.98	0.96	0.98	0.98	0.96	0.97	0.97	0.96
MZ p-value (α=0 & β=1)	0.00***	0.00***	0.52	0.00***	0.00***	0.53	0.06*	0.06*	0.14	0.00***	0.00***	0.58
HP p-value (γ=0)	0.02**	0.02**	0.56	0.05**	0.06**	0.56	0.15	0.12	0.89	0.02**	0.03**	0.57
Import Levels	**Full Sample**			**Non-Hyper**			**LIC**			**Non-LIC**		
Dependent: Final Data	**All**	**Float**	**Fixed**	**All**	**Float**	**Fixed**	**All**	**Float**	**Fixed**	**All**	**Float**	**Fixed**
Forecast (β)	0.99	1.00	0.92**	0.99	1.00	0.9**	0.86	0.86	0.96*	0.99	1.00	0.91**
p-value (β=1)	0.44	0.80	0.01	0.55	0.92	0.01	0.17	0.19	0.08	0.51	0.87	0.01
Constant (α)	-284	-327	631**	-359*	-371	776**	399	436	121	-227	-190	747**
p-value (α=0)	0.12	0.20	0.02	0.09	0.22	0.01	0.39	0.46	0.22	0.31	0.55	0.02
Observations	287	191	96	241	156	85	58	42	16	229	149	80
Adj. R^2	0.99	0.99	0.98	0.99	0.99	0.97	0.94	0.94	0.99	0.99	0.99	0.97
MZ p-value (α=0 & β=1)	0.21	0.43	0.03**	0.20	0.47	0.02**	0.02**	0.02**	0.18	0.47	0.83	0.03**
HP p-value (γ=0)	0.11	0.30	0.12	0.12	0.37	0.10	0.09*	0.09*	0.90	0.24	0.60	0.12

Robust p-values reported, *** p<0.01, ** p<0.05, * p<0.1

IV) Does IMF Reserve Forecast Accuracy Improve Over Time?

The size of our dataset allows us to examine the accuracy of Reserve Coverage forecasts not only across subsamples, but also across time. Since imports are shown to be forecast without bias for all but fixed exchange rate subsamples, we focus on Reserve Coverage and reserve levels in this section. By executing Mincer Zarnowitz regressions as in equation (1) for 5-year rolling windows across our entire sample from 1992 to 2019, we can gauge if/when bias and efficiency changed and if/when forecast coefficients and intercepts differ over time. Results are visually summarized in Figures 3a, and 3b, which reproduce the forecast coefficients and the constants, respectively, for the 5-year rolling window regressions reported in Appendix A.

Figures 3a, b succinctly summarize bias and efficiency along with the 95% confidence intervals over time. They indicate that the deep-seated bias for both Reserve Coverage and reserve levels is not just systemic across development levels, exchange rates, and trade/financial openness, but also across time. The results for the full sample are not driven by particular time periods of extreme bias and inefficiency.

It is, however, important to note that the forecasts for reserve levels start to become efficient and unbiased with the 2010-2014 window and remain so until the end of our sample in 2019 (with one inefficiency in the 2013-17 window). Reserve coverage, however, remains inefficient and biased throughout. Figures 3a and 3b hold additional, interesting implications for global crises. For the Asian Crisis, coefficients on forecasts in Figure 3a are substantially *below* unity while constants are large and positive. This indicates that during the Asian Crisis, IMF forecasts for high crisis countries (those with very low reserves in times of crisis) became especially pessimistic. In sharp contrast, we find that during the Global Financial Crisis, forecast coefficients *exceed* unity while constants were also positive. This indicates that during the Global Financial Crisis, reserve forecasts became excessively pessimistic, but increasingly so for countries with high levels of reserves. Since 2016, both reserve level and Reserve Coverage intercepts remained close to zero, indicating that the pessimistic bias for high BOP crisis countries has been mitigated.

Figure 3a: Bias and Efficiency of IMF Reserve Forecasts Over Time
Mincer Zarnowitz Forecast Coefficient (*β*)

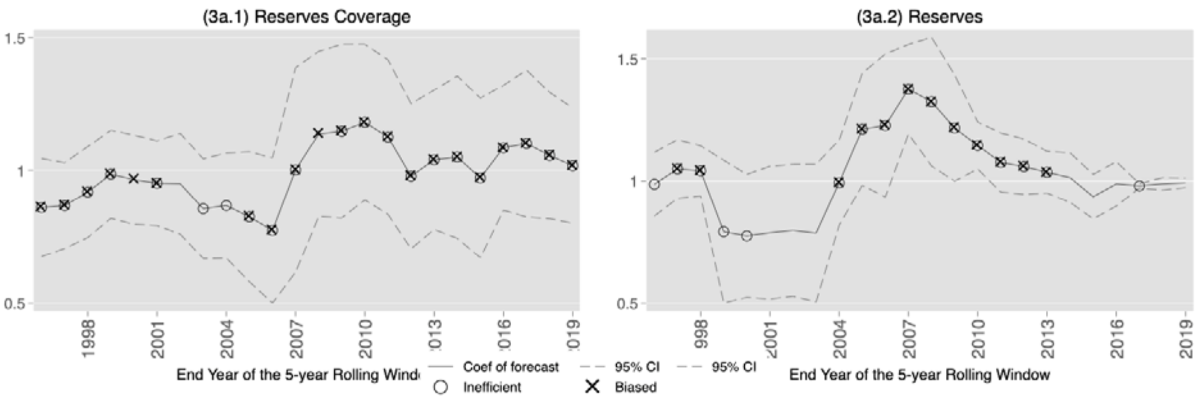
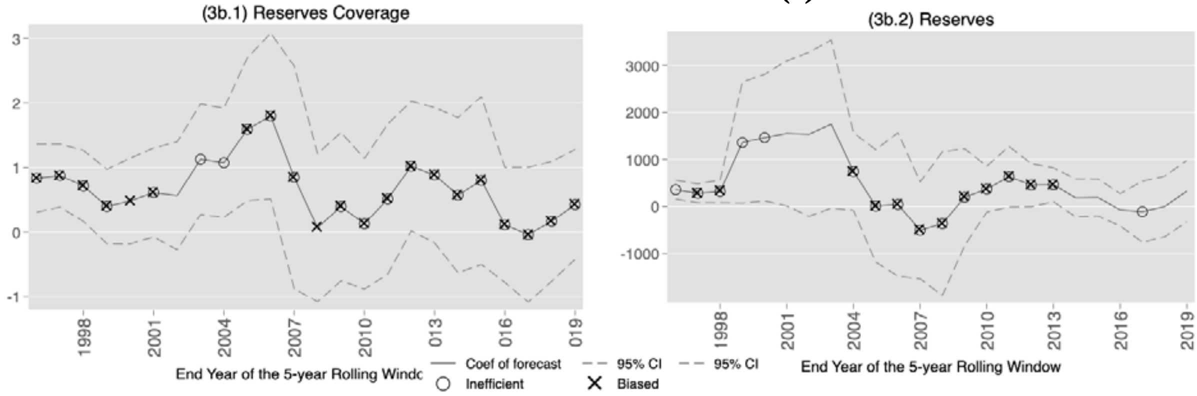


Figure 3b: Bias and Efficiency of IMF Reserve Forecasts Over Time
Mincer Zarnowitz Constant (*α*)



V) Short-Term External Debt Forecast Accuracy

Wijnholds and Kapteyn (2001) detail how the Korean STdebt crisis induced the IMF to extend the 3-month Reserve Coverage rule to include the Greenspan-Guidotti rule.[14] The rule suggests reserves should cover not only several months of imports but also 100% of STdebt to insure against sudden stops in capital flows. Foundational to IMF policy became IMF (2000), which documents that smaller ratios of reserves to STdebt are associated with greater incidences/depths of crises.

After the Global Financial Crisis, the IMF expanded the reserve metric to include the multi-faceted "Assessing Reserve Adequacy" (ARA) metric (IMF 2011). ARA captures "risk-weighted liability stocks" based on (i) "STdebt," "other liabilities which may include medium and/or long-term debt and/or equity liabilities to non-residents," (iii) Broad Money M2, and (iv) exports. STdebt retained the largest weight in ARA. At the same time, the IMF reiterated its focus on Reserve Coverage as the proper metric for countries with less open capital accounts (IMF 2015).

Since the introduction of ARA, some IMF BOP program loan documents feature reserve forecast justifications based explicitly on ARA. For researchers, however, the metric remains a black box. Program loan documents usually withhold the actual weights used to construct the country's ARA components and never provide the underlying data. MONA also does not report the ARA metric, country weights, or the underlying data. Hence we cannot

[14] Named after Argentine deputy minister of finance Pablo Guidotti and FED Chair Alan Greenspan who proposed the rule at a World Bank meeting in 1999 (see Greenspan, 1999).

assess the accuracy of ARA-based IMF reserve forecasts. Nevertheless, STdata is available, and it has been the key indicator since the early 2000s, and it remains most important ARA component. To our knowledge, we present the first and largest STdebt dataset for BOP program countries, which allows for an evaluation of STdebt forecasts.

Results in Table 4 document a similar pattern as observed for reserve forecasts: STdebt forecasts are unequivocally inefficient and often biased across subsamples. In the full sample and across subsamples that cover different income levels, exchange rate regimes, inflation, capital openness, and trade shares, we find significant inefficiencies, although the LICs sample is too small to draw meaningful inferences.

Table 4
Bias and Inefficiency of IMF Short-term External Debt Forecasts

ST Debt	ALL			Non-Hyper			LIC			Non-LIC		
Dependent: Final Data	**All**	**Float**	**Fixed**	**All**	**Float**	**Fixed**	**All**	**Float**	**Fixed**	**All**	**Float**	**Fixed**
Forecast (β)	1.03	1.09	0.51***	1.21	1.31	0.48**	1.38	1.36	7.93***	1.03	1.08	0.50***
p-value (β=1)	0.87	0.71	0.00	0.44	0.30	0.01	0.02	0.02	0.00	0.90	0.74	0.00
Constant (α)	1.64**	1.95*	1.90***	0.73	0.40	1.93***	0.22*	0.17	0.02	1.90**	2.33*	2.14***
p-value (α=0)	0.01	0.05	0.00	0.28	0.69	0.00	0.09	0.10	0.44	0.02	0.06	0.00
Observations	124	82	42	115	74	41	17	12	5	107	70	37
Adj. R^2	0.632	0.645	0.542	0.640	0.673	0.435	0.694	0.860	0.997	0.621	0.631	0.531
MZ p-value (α=0 & β=1)	0.00***	0.00***	0.00***	0.00***	0.01**	0.00***	0.01**	0.01**	0.00***	0.00***	0.01**	0.00***
HP p-value (γ=0)	0.07*	0.06*	0.83	0.06*	0.07*	0.62	0.01**	0.01**	0.35	0.07*	0.07*	0.87

ST Debt	KA High		KA Low		Trade High		Trade Low	
Dependent: Final Data	**All**	**Non-Hyper**	**All**	**Non-Hyper**	**All**	**Non-Hyper**	**All**	**Non-Hyper**
Forecast (β)	1.03	1.28	1.02	1.01	2.13*	2.46***	0.84**	0.91
p-value (β=1)	0.91	0.44	0.95	0.96	0.07	0.00	0.08	0.45
Constant (α)	2.32**	0.94	1.33*	1.03	-1.14	-1.62	1.88***	1.16***
p-value (α=0)	0.03	0.39	0.07	0.15	0.45	0.18	0.00	0.01
Observations	51	48	68	62	48	45	76	70
Adj. R^2	0.61	0.63	0.66	0.67	0.77	0.86	0.78	0.76
MZ p-value (α=0 & β=1)	0.05*	0.06*	0.00***	0.00***	0.03**	0.01**	0.00***	0.02**
HP p-value (γ=0)	0.24	0.16	0.10*	0.18	0.07*	0.07*	0.54	0.53

Robust p-values reported, *** p<0.01, ** p<0.05, * p<0.1

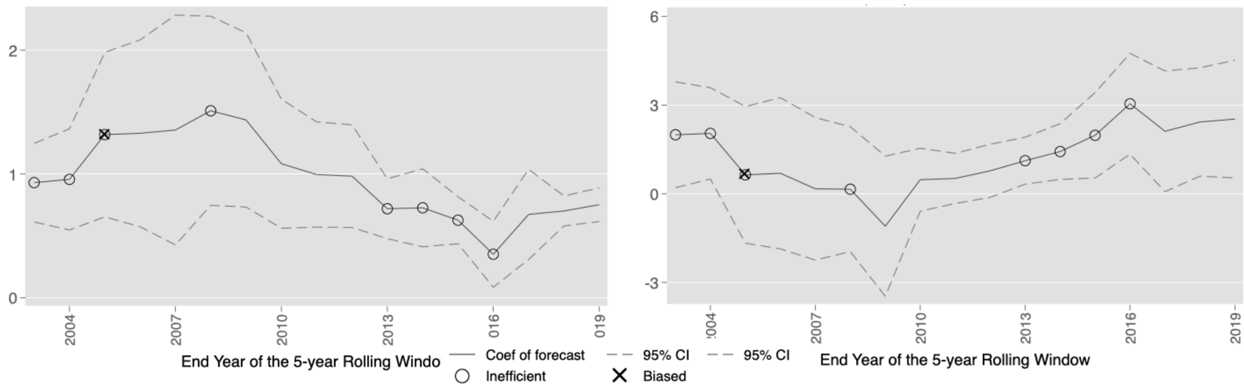
Interestingly, subsamples for fixed exchange rates, high capital mobility, and low trade shares all produce inefficient but unbiased results, but for different reasons. Fixed exchange rate subsamples all exhibit extraordinarily low coefficients on IMF STdebt forecasts paired with unusually large, positive constants (about two months Reserve Coverage). This implies that, for high crisis countries with fixed exchange rates, IMF forecasts *underestimate* STdebt along the recovery trajectory (due to the large positive constant). At the same time, IMF forecasts for low crisis countries with fixed exchange rates *overestimate* STdebt (due to the very low slope coefficient). On balance, STdebt forecasts errors for high/low BOP crises countries with fixed exchange rates offset sufficiently such that the Holden Peel (HP) cannot reject the null hypothesis that forecasts are unbiased.[15] As expected, the low slope coefficient indicates that forecasts are sufficiently inaccurate to reject efficiency. For high capital mobility and low trade shares, the coefficients on IMF forecasts are around unity but for high trade shares, the coefficient exceeds 2, meaning that, on average, IMF forecasts are too optimistic and underestimate STdebt by up to 100% relative to the observed outcomes. Aside from fixed exchange rate subsamples, all other samples exhibit positive constants and coefficients for IMF forecast that equal or exceed unity. This implies that IMF STdebt forecasts have a significant optimistic bias in the sense that IMF STdebt forecasts for program countries are significantly lower than the subsequently observed outcomes.

Table 4 also suggests an interesting hypothesis for the biased and inefficient reserve forecasts that we observed for reserves: excessively optimistic (low) STdebt forecasts may cause excessively pessimistic (low) levels of reserves. But when STdebt outcomes turn out to be routinely higher than forecast, it also requires higher reserves to cover the surprise increase in short-term liabilities. While it makes for a good story, the hypothesis is unlikely to apply to our data, since we are using forecasts from crisis year *t* for year *t.* It is unlikely that significant, systematic debt *and* reserve revisions occur during the first months of the program.

Of interest is again whether forecast accuracy improved over time. Figure 4 plots the evolution of bias and inefficiency from 2003 to 2019. Before 2003, rolling windows contain less than 20 observations, hence we report only results from 2003 onward. The 5-year windows indicate most of the forecast inefficiencies are generated in recent years, after 2013, when coefficients on forecasts decline significantly below unity while constants move into significant, and hugely positive territory. This is a remarkably different result as compared to our findings for reserves, where reserve levels became less inaccurate in recent years.

[15] Recall that bias is a weaker concept than efficiency as wildly inaccurate estimates, such as a 40% overestimate and 40% underestimate average to zero bias.

Figure 4
The Evolution of Bias and Inefficiency Over Time
IMF Short-term External Debt Forecasts



VI) Program Cancellation, Conditionality, Loan Size, and Geopolitics as Drivers of Reserve and Short-Term External Debt Forecast Inefficiencies

VI.1) Program Cancellations

Atoyan and Conway (2011), Luna (2014), and IMF (2019) point out that IMF forecasts are conditional on countries' proper implementation of program conditions; hence, implementation failures may explain IMF forecast bias. Note, however, that implementation failures can explain the optimistic STdebt bias, but not the pessimistic reserve bias we observed above. Implementation failures along with exogenous, non-economic events (e.g., natural disasters, civil wars and international conflicts) may cause IMF program cancellations. In this section, we ask if exogenous shocks and program cancellations may have driven biased IMF forecasts.

Our dataset contains 47 cancelled programs. In Table 5, we report baseline results without canceled programs for Reserve Coverage, reserve levels, import levels, and STdebt forecasts. Results remain qualitatively unchanged. Even without cancelled programs, reserves, Reserve Coverage, and STdebt are biased and inefficient while imports remain just as unbiased and efficiently forecast as in the sample that included cancellations. Table 5 thus indicates program cancellations cannot explain forecast inaccuracies. This finding is similar to Phillips and Musso (2002) who included a "program interruption dummy," which never turned significant.

Table 5: Bias and Inefficiency of IMF Forecasts
Cancelled Programs Excluded

Dependent: Final Data	Reserve Coverage	Reserve Level	Imports	ST Debt
Forecast (*β*)	0.97	1.02	0.98	1.04
p-value ($\beta=1$)	0.62	0.27	0.27	0.80
Constant (*α*)	0.65***	482.61***	-46.29	1.54***
p-value ($\alpha=0$)	0.00	0.00	0.73	0.00
Observations	240	240	240	102
Adj. R^2	0.75	0.98	0.99	0.73
MZ p-value ($\alpha=0$ & $\beta=1$)	0.00***	0.00***	0.53	0.00***
HP p-value ($\gamma=0$)	0.00***	0.00***	0.29	0.01**

Robust p-values reported, *** p<0.01, ** p<0.05, * p<0.1

VI.2) Conditionality, Loan Size, and Geopolitics

If program cancellations did not affect the inaccuracies of IMF forecasts, we aim to cast a wider net in this section and examine *which* information that was available to forecasters at the time of the forecasts, may not have been properly integrated into IMF forecasts. Aside from program cancellation, we study three additional categories of information that relate to (i) IMF conditionality, (i) loan size, and (iii) geopolitical events. First, we lay out the empirical methodology and then we discuss our findings examining sources of forecast inefficiencies for Reserve Coverage, reserve levels and STdebt (we exclude imports since their forecasts are unbiased and efficient).

We follow the approach proposed by Eicher and Gao (2022), who examine the effects of all types of conditions on IMF GDP and inflation forecasts. Eicher and Kawai (2022) follow that same approach for IMF import, export and exchange rate forecasts. Instead of selecting to include a particular subset of conditionality, we include the entire range of 11 thematic categories for conditionality that has been established by the IMF in the MONA database.[16]

There are additional areas that have previously been linked to IMF forecast errors. Beach et al. (1999), Dreher et al. (2008), and Luna, (2014) all show that the size of the IMF loans affects IMF forecast accuracy. Hence we also investigate if the program loan-to-quota ratio affects forecast accuracy. In essence, we are asking if there is a systematic optimistic or pessimistic bias in the reserve and debt forecasts for larger programs. The third area that has been suggested as exerting an effect on IMF forecasts relates to economic effects of

[16] The thematic areas are Total Domestic Credit, Government/Public Sector Credit, BOP/Reserve Test, Medium/Long-Term External Debt, Subceiling on Medium/Long-Term External Debt; STdebt, No New Arrears/Defaults (Continuous Injunction), Ceilings on External Arrears, Fiscal Deficit, Domestic Arrears.

geopolitical events that are known to IMF forecasters at the time of forecasts. For example, the IMF's *Review of IMF Program Design and Conditionality* (IMF, 2019) notes that forecast errors are impacted by political transitions, conflicts, and natural disasters (see also Przeworski and Vreeland, 2000; Park, 2006; Mody and Rebucci, 2006; Kentikelenis et al, 2016). Hence, we consider variables related to elections (executive and legislative) up to one year before the start of a program. For conflicts, we include indicators for civil wars and international conflicts that commenced up to one year prior to the start of a program. Finally, we consider natural disasters that occurred up to one year prior to the start of a program (see Appendix B for election, conflict, and disaster data). Crucial is that all geopolitical events were known at the time of forecasts, so IMF forecasters were well aware of their potential effects on economic performance. In addition, we allow for "exogenous" election and disaster events to understand if, perhaps, forecasts were systematically biased by events that occurred soon after the program commenced and after the forecasts were established.

Sinclair et al. (2010) extended Mincer and Zarnowitz's (1969) to analyze bias and inefficiency in the presence of potentially imperfectly integrated information, captured by a vector of candidate covariates, X_t. These covariates represent information available to forecasters at the time of the forecast that was perhaps not been fully integrated into the forecasts:

$$A_t = \alpha + \beta F_t + \delta X_t + \varepsilon_t, \tag{2}$$

Any non-zero δ indicates that the information contained in the candidate covariate is systematically not fully integrated into forecasts to contribute to bias and inefficiency. Statistically significant covariates thus represent areas that the IMF might consider with particular interest in future IMF program forecasts. Sinclair et al. (2010) propose the joint null hypothesis of $\beta=1$ & $\alpha = \delta = 0$ as a formal test of whether the information contained in the additional covariates was properly included in the forecast. If the null is rejected, Sinclair et al. (2010) cite the test statistic as evidence that information contained in X was not fully integrated into the forecast, to pose a source of bias and inefficiency.

We apply the Sinclair et al. (2010) methodology to our three categories of interest that were identified above as possible drivers of IMF forecast errors in Table 6. Given the sizable number of covariates, Table 6 reports only results for significant regressors to economize on space. The Sinclair et al. (SJS, 2010) tests in Table 6 indicate that a number of additional, significant regressors cannot be rejected to have contributed to bias and inefficiency for

Reserve Coverage, reserve levels, and STdebt forecasts. This implies that information known at the time of the forecast was not properly integrated for all three measures.

Program size is indeed one factor that could have improved forecasts, but only for reserve levels. Interestingly the regression in Table 6 indicates that, the greater the loan size, the larger the overestimate of Reserve Coverage. Global crises also exert effects that are not fully captured by IMF forecasts. During the 2008 Global Financial Crisis (programs that started *after* the crisis commenced on Sept 2008), STdebt forecasts were systematically higher. Perhaps another case of including a risk buffer during times of extreme uncertainty. In terms of the Asian Crisis, we find that countries that started programs *before* the crisis commenced were subject to Reserve Coverage forecasts that were too optimistic. This is no surprise since the crisis was exogenous to the forecast as it occurred after the program forecast had been finalized.

In terms of conditionality, we find that, when BOP reserve tests are part of IMF conditionality, IMF Reserve Coverage and reserve level forecasts are particularly pessimistic (significantly too low). Reserve Tests involve conditionality that reflects a floor in net international reserves. Table 6 thus suggests that countries with net international reserve floor conditionality systematically receive pessimistic gross international reserve forecasts. Reserve Coverage is always based on gross reserves, so our finding indicates special attention must be given to gross reserve forecasts whenever net reserve floors are also specified for a program.

Conditionality limiting government credit is also not properly integrated into reserve level forecasts, leading to an overestimate of future reserve levels. When conditionality imposes limits arrears (external or default limits), STdebt is consistently forecasted too high. One may surmise whether an (unnecessary) buffer is systematically built into programs when arrears are limited but the program underestimates STdebt which increases to adhere to the arrears conditionality. When conditionality limits external arrears, STdebt is consistently pessimistically estimated.

Table 6
Sources of Forecast Bias and Inefficiency of IMF Forecasts
Reserve Coverage, Reserve Levels, Short-Term External Debt

	Dependent: Actual Final Data	Reserve Coverage	Reserve Levels	S-T Debt
	Forecast	0.92	0.98	1.00
	p-value (*β=1*)	0.12	0.44	0.99
	Program Size ($/Quota)	-1.39	31,939.20**	40.57
		0.78	0.02	0.59
Crises	**2008, endogenous**	0.31	-309.40	-4.31*
		0.49	0.67	0.09
	1997, exogenous	-1.09*	-233.94	na
		0.08	0.88	na
Conditionality	**BOP Reserve Test**	0.43*	1,121.59**	-1.04
		0.09	0.01	0.70
	Arrears (No New)	-0.22	417.75	-5.23*
		0.31	0.48	0.08
	Arrears (Ceiling on Ext.)	-0.24	-61.16	-6.67**
		0.25	0.88	0.01
	Fiscal (Credit to Gov't)	-0.30	-1,513.34**	-0.15
		0.23	0.02	0.93
Non-Economic	**Nat Disaster, endogenous**	-0.52**	-900.74*	-4.44
		0.02	0.06	0.14
	Wars, exogenous	-1.39**	-2,358.32	
		0.03	0.32	
	Civil Wars, endogenous	0.69**	977.13	4.44*
		0.01	0.11	0.06
	Constant	1.06**	279.19	2.26
	p-value (*α=0*)	0.03	0.67	0.48
	Observations	286	287	124
	Adjusted R-squared	0.74	0.98	0.64
	SJS p-val $\beta=1$ & $\alpha=\delta=0$	0.00***	0.00***	0.03**

Robust p-values reported, *** p<0.01, ** p<0.05, * p<0.1 Only significant regressors reported, other regressors included in the regression include endogenous and exogenous crises, elections, civil wars, international conflicts ("wars"), disasters, trade share, and all types of conditionality included in MONA (see footnote 15).

Endogenous natural disasters, known to have occurred at the time of the forecast, are not properly taken into account for Reserve Coverage and reserve level forecasts leading to systematically optimistic forecasts. Finally, international conflicts and civil wars are not properly accounted for in Reserve Coverage and STdebt. As expected, exogenous wars that occur after the forecast and the program have been finalized, lead to an overestimate of Reserve

Coverage. Civil wars that are known at the time of the forecast, on the other hand, lead to an underestimate of both reserves and STdebt. The latter is intuitive, the former is a bit of a puzzle.

VII) Robustness

VII.1) Time Horizons and Forecast Accuracy: *t+1* and *t+2* Forecasts

Above we examine forecasts produced in year *t* for year *t.* Just like predicting the weather tomorrow is easier than the weather next month, it is generally believed longer forecast horizons produce lower forecast accuracy. Longer forecast horizons increase uncertainties, shocks, and the number of factors that may affect the future to render the original data and assumptions outdated or irrelevant (see Armstrong, 2001 and USGAO, 2003). Timmermann (2007) found IMF forecast errors increase with time horizons (he used WEO data that did not include crisis countries).

On the other hand, one might argue that for IMF programs, longer horizons may allow more time for program success to manifest itself as policy adjustments take time to be implemented and affect the economy. For this reason, Blanchard and Leigh (2013) and Carrière-Swallow and Marzluf (2022) consider two-year forecast horizons. Of course, one hopes the understanding that policy changes take time to be implemented would also be factored into short-term forecasts (in *t* for *t*). However, it is certainly of interest to explore if policy adjustments and program success outweigh additional uncertainties to produce potentially increased forecast accuracy over longer time horizons.

To gain an understanding of whether forecast accuracy increase or decreases with time horizons, we examine forecasts (produced in *t)* for years *t+1* and *t+2.* Results for longer time horizons are reported in Tables 1a and 1b below.[17] We find that bias and inefficiency remain unchanged as the forecast horizon expands. Only LICs with fixed exchange rates were unbiased and efficient in Table 1, likely due to the lower power produced by the 15 observations. The same pattern is reproduced in Tables 1a and 1b for forecasts for years *t+1* and *t+2.*

While longer forecast horizons do not produce different bias and inefficiency across subsamples, but there are two important insights that we can obtain from the exercise. The forecasts become successively more inaccurate as the time horizon expands. First, the bias

[17] Results for the other tables are qualitatively similar and available upon request. The number of observations varies slightly as not all programs report longer horizon forecasts in MONA or in loan documents.

increases as the constant increases substantially from *t* to *t+1* and then *t+2* in just about all samples. This indicates that the forecast error for the most vulnerable crisis countries increases with the time horizon. Secondly, the slope coefficient measuring forecast efficiency becomes successively smaller for all subsamples as the forecast horizon is extended to *t+1* and then t+2. Comparing Tables 1a/1b to Table 1 suggest that successively longer forecast horizons, imply that the IMF builds ever greater caution into forecasts.

Table 1a
Bias and Inefficiency of IMF Reserve Coverage Forecasts in *t* for *t+1*

Reserve Coverage	ALL			Non-Hyper			LIC			Non-LIC		
Dependent: Final Data	**All**	**All Float**	**All Fixed**	**All**	**All Float**	**All Fixed**	**All**	**All Float**	**All Fixed**	**All**	**All Float**	**All Fixed**
Forecast (β)	0.87***	0.89***	0.82***	0.88***	0.91***	0.81***	0.97***	1.24***	0.81***	0.85***	0.85***	0.82***
p-value (β=1)	0.008	0.052	0.070	0.020	0.091	0.094	0.831	0.049	0.092	0.008	0.015	0.208
Constant (α)	1.30***	1.28***	1.35***	1.29***	1.20***	1.51***	1.01**	0.190	1.55*	1.36***	1.43***	1.29**
p-value (α=0)	0.000	0.000	0.004	0.000	0.000	0.005	0.019	0.659	0.055	0.000	0.000	0.037
Observations	270	181	89	225	147	78	54	39	15	216	142	74
Adj. R^2	0.616	0.661	0.523	0.635	0.701	0.508	0.718	0.737	0.755	0.589	0.647	0.454
MZ p-value (α=0 & β=1)	0.0***	0.0***	0.0***	0.0***	0.0***	0.0***	0.0***	0.0***	0.149	0.0***	0.0***	0.02**
HP p-value (γ=0)	0.0***	0.0***	0.0***	0.0***	0.0***	0.0***	0.0***	0.0***	0.097	0.0***	0.0***	0.03***

Robust p-values reported, *** p<0.01, ** p<0.05, * p<0.1

Table 1b
Bias and Inefficiency of IMF Reserve Coverage Forecasts in *t* for *t+2*

Reserve Coverage	ALL			Non-Hyper			LIC			Non-LIC		
Dependent: Final Data	**All**	**All Float**	**All Fixed**	**All**	**All Float**	**All Fixed**	**All**	**All Float**	**All Fixed**	**All**	**All Float**	**All Fixed**
Forecast (β)	0.83***	0.84***	0.82***	0.83***	0.85***	0.78***	0.94***	1.06***	0.90***	0.81***	0.82***	0.79***
p-value (β=1)	0.008	0.039	0.104	0.014	0.093	0.067	0.534	0.704	0.383	0.010	0.031	0.160
Constant (α)	1.31***	1.34***	1.24**	1.34***	1.29***	1.45***	0.94**	0.580	1.070	1.40***	1.44***	1.38**
p-value (α=0)	0.000	0.000	0.011	0.000	0.001	0.009	0.013	0.250	0.151	0.000	0.000	0.042
Observations	241	161	80	199	129	70	50	35	15	191	126	65
Adj. R^2	0.528	0.546	0.485	0.559	0.616	0.457	0.681	0.602	0.720	0.489	0.525	0.390
MZ p-value (α=0 & β=1)	0.0***	0.0***	0.02**	0.0***	0.0***	0.02**	0.0***	0.0***	0.335	0.0***	0.0***	0.059**
HP p-value (γ=0)	0.0***	0.0***	0.059*	0.0***	0.0***	0.08*	0.0***	0.0***	0.196	0.0***	0.0***	0.137

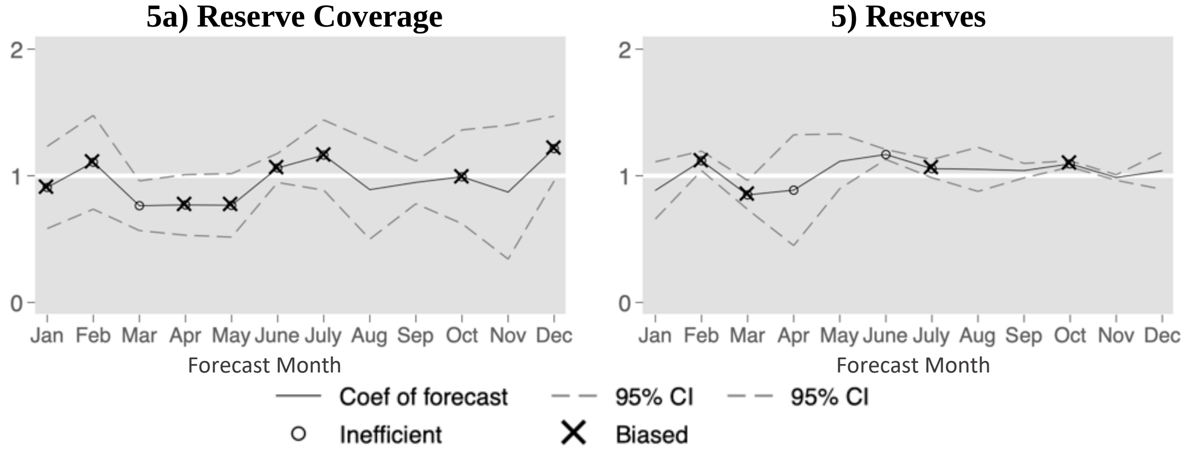
Robust p-values reported, *** p<0.01, ** p<0.05, * p<0.1

VII.2) Time Horizons and Forecast Accuracy: Calendar Dates

A similar exercise related to the forecast horizon can be performed for forecasts produced in year *t* for year *t*. One might suspect that for the short-term time horizon, the month in which the forecast was established might matter greatly. Forecasters' information sets increase substantially in size and accuracy towards later months of the year, as compared to the data vintages available in early months. Below we examine whether forecast horizons as measured by calendar months' drive bias and inefficiency in the short-term.

Figures 5a) and 5b) provide a visual summary of the change in Reserve Coverage and reserve forecast accuracies as the time horizon in year *t* changes. The table reports results from Mincer-Zarnowitz regressions and the vertical axis displays regression coefficients, while the horizontal axis displays the month in which the forecast was made (see Appendix Table A.4 for the raw regression results). The figures also identify biased and inefficient forecasts according to the Figures' legend. As in Eicher and Gao Rollinson (2020) and Eicher and Kawai (2022), who examined IMF forecasts for GDP, inflation, imports, and exchange rates in crisis countries, we do not find that bias and efficiency of reserve or Reserve Coverage forecasts improve in later months during the calendar year when presumably more data and less uncertainty is present. Biases and inefficiencies are distributed roughly evenly across the year without a clear pattern of either bias or efficiency improvements as forecast horizons change.

Figure 5: Are Earlier Forecasts in *t* More Accurate?



VII.3) Does The Source of Final Outcome Data Matter?

Accurate forecast evaluations require accurate corresponding mapping to final outcome data. Above we evaluate MONA forecast for gross reserves that are produced at the start of IMF programs by using actual gross reserve outcome data from WEO. A referee suggested that gross reserve data in levels, as reported in IMF loan documents (and recorded by MONA) may not always coincide with the IMF's Balance of Payments Manuals' definition of gross reserves. We studied the loan documents and did not find evidence of programs redefining gross reserves, nor that IMF BOP Manuals changed definitions since the 1990s. Nevertheless, it is of interest to examine whether bias and inefficiency of IMF reserve forecasts differ when we use MONA's vintage of final outcome data to assess the accuracy of forecasts.[18]

MONA final outcome data is defined as data for time *t*, reported at the last recorded program review. Since the last IMF program review always takes place at least a year after programs start, this MONA data vintage can be seen as a proxy for actual final outcome data (see also Eicher et al, 2019). There is room for disparities between MONA and WEO final data, given that data revisions occur after IMF programs conclude (i.e., Eicher Gao Rollinson, 2022). In addition, we document substantial MONA database errors and missing observations in Appendix C. WEO final reserve data allows us to evaluate gross reserve forecasts for 287 IMF programs, while MONA final data allows us to evaluate only 122 forecasts after accounting for data cleaning and missing data. Comparing WEO to MONA final reserve data, 25 percent of the observations are identical; 40 percent fall within 2% of each other, and the correlation is 0.98.

Table 3b reports results for Table 3a produced using the substantially smaller MONA final data sample. It highlights a pattern of bias and inefficiency in the global sample and across subsamples that are just about identical for MONA (Table 3b) and WEO (Table 3a) data. The one noticeable difference is that LICs are no longer biased and inefficient, but this is almost certainly due to the fact that the LICs MONA samples have fewer than 25 observations. The small number of LICs compromises the power of regression, raising questions of how meaningful LICs results are when using MONA final outcome data.

We also observe differences in the deviations of regression lines from the Line of Perfect Forecast. The intercept for the regressions based on MONA final data increased dramatically, almost doubling in several cases. This suggests increased bias, where actual final reserves reported in MONA are significantly larger compared to WEO data for high-crisis countries (those with below-average reserves at the start of the program). One explanation may

[18] Net Reserves are, at times. individually defined in specific IMF programs (hence no uniform definition for Net Reserves exist in the IMF BOP manuals); Gross Reserves (the data we use above), however, follow one uniform definition in the IMF BOP manuals dating back to the 1980s.

be that reserve data in IMF final program reports (as recorded in MONA) are systematically revised downward after reports have been submitted.

The increase in the intercept produced by MONA data in Table 3b is matched by lower slope coefficients in the MONA regressions. The lower slope coefficients in Table 3b compared to Table 3a indicate a decline in efficiency as more information known at the time of forecast was ignored in the forecast sample based exclusively on MONA data. Table 3b suggests that, for the full sample, the average forecast based on MONA final data is even more skewed towards over-optimistic reserve forecasts than using WEO data.

Table 3b: Does The Source of Final Outcome Data Matter for Bias and Inefficiency? MONA Data as Final Outcome Data

Reserves Levels	**ALL**			**Non-Hyper**			**LIC**			**Non-LIC**		
Dependent: Final Data	**All**	**All Float**	**All Fixed**	**All**	**All Float**	**All Fixed**	**All**	**All Float**	**All Fixed**	**All**	**All Float**	**All Fixed**
Forecast (β)	0.83***	0.76***	1.05***	0.79***	0.70***	1.05***	0.91***	0.90***	1.00***	0.82***	0.72***	1.05***
p-value ($\beta=1$)	0.20	0.09	0.23	0.13	0.03	0.25	0.42	0.44	0.09	0.22	0.08	0.25
Constant (α)	855.71*	1,283.44**	6.91	911.12*	1,261.57**	29.01	120.11	102.46	21.20*	1,046.51*	1,570.09**	20.80
p-value ($\alpha=0$)	0.09	0.04	0.93	0.08	0.02	0.78	0.42	0.55	0.08	0.08	0.02	0.84
Observations	122	77	45	90	53	37	25	15	10	97	62	35
Adj. R^2	0.88	0.84	0.99	0.88	0.85	0.99	0.95	0.93	1.00	0.86	0.82	0.99
MZ p-value ($\alpha=0$ & $\beta=1$)	0.12	0.09*	0.15	0.20	0.06*	0.09*	0.70	0.73	0.19	0.08*	0.05*	0.15
HP p-value ($\gamma=0$)	0.02**	0.02**	0.56	0.05*	0.06*	0.56	0.15	0.12	0.89	0.02**	0.03**	0.57

Robust p-values reported, *** p<0.01, ** p<0.05, * p<0.1

VII.4) Variable Definition: Does Forecast Inaccuracy Differ for Levels Vs Growth Rates?

If data revisions matter to our analysis (they may in terms of the suggested size of the bias and inefficiency), it may be of interest to examine forecasts not only in levels but also in growth rates. Recall that we use the level of Reserve Coverage because IMF loan documents report this quantity, the program is designed around the level and the rule of thumb of adequate reserves per months imports is based on the level. However, it may be that changes from (*t-1 to t*) are less likely affected by large ex-post data revisions. Certainly, growth rates and especially log changes compress the range or scale of the data. This approach then deemphasizes the impact of country crises that exhibited unusually large/small Reserve Coverages levels. Table 1c uses log changes of Reserve Coverage while Table 1 used levels.

Results are again very similar and the bias and inefficiency patterns across subsamples are confirmed for all but fixed exchange rate subsamples. The log change approach suggests forecasts for fixed exchange rate LICs are biased and inefficient – although this result is based on far too few observations to instill confidence. Interestingly, fixed exchange rates in the full sample, as well as the Non-Hyperinflation and Non-LICs samples are not biased and inefficient in Table 1c. Aside from contrasting bias and inefficiency, we refrain from comparing the size of intercepts and coefficient in level vs growth rate regressions since they carry very different interpretations.

Table 1c: Bias and Inefficiency When Forecast Evaluation Is Based on Log Changes

Reserve Coverage	**ALL**			**Non-Hyper**			**LIC**			**Non-LIC**		
Dependent: Final Data	**All**	**All Float**	**All Fixed**	**All**	**All Float**	**All Fixed**	**All**	**All Float**	**All Fixed**	**All**	**All Float**	**All Fixed**
Forecast (*β*)	0.84***	0.67***	1.09***	0.71***	0.67***	0.83***	0.59***	0.36***	1.51***	0.91***	0.80***	1.04***
p-value (β=*1*)	0.16	0.00	0.55	0.00	0.00	0.50	0.05	0.00	0.00	0.45	0.02	0.84
Constant (*α*)	0.07***	0.08***	0.05	0.07***	0.07***	0.06*	0.10**	0.05	0.21***	0.06***	0.09***	0.01
p-value (α=*0*)	0.00	0.00	0.18	0.00	0.00	0.09	0.01	0.32	0.01	0.00	0.00	0.72
Observations	281	187	94	236	153	83	56	41	15	225	146	79
Adj. R^2	0.49	0.42	0.59	0.38	0.40	0.35	0.29	0.18	0.81	0.54	0.51	0.58
MZ p-value (*α=0 & β=1*)	0.00***	0.00***	0.32	0.00***	0.00***	0.24	0.00***	0.00***	0.00***	0.01**	0.00***	0.91
HP p-value (*γ=0*)	0.00***	0.02**	0.13	0.01**	0.02**	0.16	0.20	0.95	0.00***	0.01**	0.00***	0.66

Robust p-values reported, *** p<0.01, ** p<0.05, * p<0.1

VIII) Conclusion

An active literature has been evaluating the accuracy of IMF forecasts for program countries over the past 35 years. This literature focuses on prominent macroeconomic aggregates: GDP, inflation, fiscal deficits, and even national accounts. For countries that experience balance of payment crises, the evaluation of external account forecasts (imports, exports, exchange) has been a central focus. Surprisingly, two key indicators that anchor IMF BOP crisis program trajectories have not been carefully studied: reserves and short-term external debt.

We audit the IMF MONA database, correct errors, and fill in missing data using IMF archival loan documents (the original source of the MONA data). Our resulting dataset has nearly 300 observations over 28 years, which allows for the first statistical analysis of forecast accuracy by subsamples covering exchange rate regimes, degrees of capital mobility, trade shares, levels of development, and across time.

We find IMF reserve forecasts for BOP program countries are characterized by systemic pessimistic bias and inefficiency, a result that is strikingly robust across all subsamples. IMF program forecasts systematically underestimate reserves countries accumulate in the year the program commences and this pessimism increases with the level of BOP crisis. Countries with the least BOP reserves at the start of their crises find that reserve forecasts are on average 55% lower than the actual observed outcomes. We show that the bias and inefficiency of Reserve Coverage forecasts are almost entirely driven by bias and inefficiencies in forecasts for reserve levels, not imports.

Following the Asian Financial Crisis, the IMF also included a focus on Short-Term external debt (STdebt) as an indicator to peg the appropriate level of reserves. STdebt forecasts are shown to be just as biased and inefficient, but unfortunately with a significant optimistic bias. IMF programs project systematically lower STdebt levels than what is eventually observed in the final data. Perhaps the systematically low STdebt forecasts may drive the errors in the systematically higher-than-forecasts reserves that we observe in the final data. Examining the accuracy of reserves and debt over time, we find that since approximately 2014, debt and reserve forecasts have improved remarkably to become unbiased and efficient. We speculate this may be a positive byproduct of the use of the new multi-indicator IMF reserve adequacy metric.

Aside from STdebt as a driver of reserve forecast inaccuracies, we also examine whether the effects of IMF conditionality and noneconomic events were properly integrated into IMF forecasts. Most importantly, conditionality related to BOP reserve tests is not properly integrated into IMF forecasts to drive inefficiency. Reserve and/or debt forecast inaccuracy is shown to be driven by conditions on fiscal deficits and arrears. The connection between these conditions and the bias and inefficiency of IMF reserve and debt forecasts is left for future research.

References

Aizenman, J and H. Genberg. (2012). "Research on the Demand for International Reserves: Developments in Academia, the Contribution of IMF Researchers, and Influence on IMF Surveillance" IMF IEO Background Paper BP/12/01

Armstrong, J.S. ed. (2001). *The Principles of Forecasting*. Norwell, Mass.: Kluwer Academic Publisher.

Artis, M. J. (1988). "How Accurate is the World Economic Outlook? A Post Mortem on Short-term Forecasting at the International Monetary Fund." *IMF Staff Studies for the IMF WEO*, Washington, DC: IMF.

Artis, M. J. (1996). "How Accurate are the WEO's Short-term Forecasts? An Examination of the World Economic Outlook." *IMF Working Paper* No. 96/89, Washington, DC: IMF.

Atoyan, R., and Conway, P. (2011). "Projecting macroeconomic outcomes: evidence from the IMF." *Review of International Organizations 6*, (3), pp. 415-441.

Baker, D. and Rosnick, D., 2003. Too sunny in Latin America? The IMF's overly optimistic growth projections and their consequences. *NACLA Report on the Americas*, *37*(3), pp.6-7.

Baqir, R., Ramcharan, R., and Sahay, R. (2006). "IMF programs and growth: Is optimism defensible?" In Mody, M. A. and Rebucci M.A. eds *IMF-supported programs: recent staff research*. Washington DC.: IMF.

Barrionuevo, J. M. (1993). "How accurate are the WEO projections?" *Staff Studies for the World Economic Outlook*, Washington, DC: IMF.

Beck, T., Clarke, G., Groff, A., Keefer, P., and Walsh, P. (2001). "New Tools in Comparative Political Economy: The Database of Political Institutions." *The World Bank Economic Review*, *15*(1), pp.165–176.

Beach, W., Schavey, A., and Isidro, I. (1999). "How reliable are IMF economic forecasts?" *Heritage Foundation Center for Data Analysis Report 99-05*, Washington D.C.: Heritage Foundation.

Beaudry, P. and Willems, T. (forthcoming). "On macroeconomic consequences of over-optimism." *American Economic Journal: Macroeconomics.*

Blanchard O. J. and Leigh D. (2013). "Growth Forecast Errors and Fiscal Multipliers," *American Economic Review*, 103(3), pp 117-20.

Bordo, M, A. Mody and Oomes N. (2004). "Keeping Capital Flowing: The Role of the IMF" IMF WP/04/197

Carrière-Swallow, Y. and Marzluf, J., 2022. Macrofinancial Causes of Optimism in Growth Forecasts. *IMF Economic Review*, pp.1-29.

Celasun, O., Lee, J., Mrkaic, M. M. and Timmermann, M. A., 2021. *An Evaluation of World Economic Outlook Growth Forecasts, 2004–17*. Washington DC.: IMF.

Chinn, M. D. and Ito, H. (2006). What matters for financial development? Capital controls, institutions, and interactions. *Journal of Development Economics*, *81*(1), pp. 163-192.

Chitu, L. J. Gomes and R. Pauli. (2019). "Trends in Central Banks' Foreign Currency Reserves and the Case of the ECB." ECB Economic Bulletin, 7/2019

Dreher, A., Marchesi, S and Vreeland, J. (2008). "The political economy of IMF forecasts." *Public Choice*, 137, (1-2), pp. 145-171

Dornbusch R. and Fischer S. (1986). "Stopping hyperinflations past and present." *Weltwirtschaftliches Archiv*, 122 (1) (1986), pp. 1-47

Eicher, T. S., and Gao Rollinson, Y. (2022). "The accuracy of IMF crises nowcasts." *International Journal of Forecasting*, forthcoming.

Eicher, T.S. and Kawai R. (2022). "IMF trade forecasts for crisis countries: Bias, inefficiency, and their origins," *International Journal of Forecasting*, forthcoming.

Eicher, T. S., Kuenzel, D.J., Papageorgiou, C., and Christofides, C. (2019). "Forecasts in times of crises." *International Journal of Forecasting*, 35, pp. 1143–1159.

EM-DAT (2020). *The OFDA/CRED International Disaster Database*. Date accessed: 10/30/20 http://www.emdat.be

Fernandez D, Ley E, and Steel M. (2001). "Benchmark Priors for Bayesian Model Averaging." *Journal of Econometrics*, 100, issue 2, pp. 381-427

Ferrer J. and Kireyev A. (2022). "Short-term Response to a Catastrophic Event" IMF Working Paper WP/22/123

Genberg, H. and Martinez, A. (2014). "On the Accuracy and Efficiency of IMF Forecasts: A Survey and Some Extensions." *Independent Evaluation Office BP/14/04*, Washington DC: IMF.

Greenspan, A. (1999). "Currency Reserve and Debt." Remarks Before the World Bank Conference on *Recent Trends in Reserves Management*, Washington, D.C.

Holden, K., and Peel, D. A. (1990). "On testing for unbiasedness and efficiency of forecasts." *The Manchester School*, 58, (2), pp. 120-127.

Holden, K., Peel, D., and Sandhu B. (1987). "The accuracy of OECD forecasts." *Empirical Economics*, 12, pp 175–186.

Harbom, L., Strand, H., Nygård, H. M. (2009). *UCDP/PRIO Armed Conflict Dataset Codebook. Version 20.1*. Date Accessed: 10/30/20. URL https://www.ucdp.uu.se/database

IFES (2020). *ElectionGuide by the International Foundation for Electoral Systems (IFES).* Date accessed: 10/15/20 URL https://www.electionguide.org/elections

Ilzetzki, E., Reinhart, C. M. and Rogoff, K. S. (2019). "Exchange rate arrangements entering the 21st century: which anchor will hold?" *Quarterly Journal of Economics*, Oxford University Press, 134, (2), pp. 599-646.

IMF. (2000). "*Debt- and Reserve-Related Indicators of External Vulnerability*" IMF Working Paper SM/00/65.

IMF. (2011). "*Assessing Reserve Adequacy*." IMF Policy Paper.

IMF. (2013). "*Assessing Reserve Adequacy – Further Considerations.*" IMF Policy Paper.

IMF. (2015). "*Assessing Reserve Adequacy – Specific Proposals.*" IMF Staff Report

IMF. (2016). "*Guidance Note on the Assessment of Reserve Adequacy and Related Considerations.*" Staff Report

IMF. (2019). "*2018 Review of Program Design and Conditionality."* Washington DC: IMF.

IMF. (2020). "*Eligibility to Use the Fund's Facilities for Concessional Financing."* Washington DC: IMF.

IMF. (2021a). "*Monitoring of Fund Arrangements."* Washington DC: IMF.

IMF. (2021b). "*World Economic Outlook."* Washington DC: IMF.

IMF. (2021c). "*International Financial Statistics."* Washington DC: IMF.

IMF. (2021d). "*Balance of Payments Statistics Yearbook (Balance of Payments and International Investment Position)."* Washington DC: IMF.

IMF. (2021e). "*Annual Report on Exchange Arrangements and Exchange Restrictions."* Washington DC: IMF.

IMF. (2022b). "Lessons from Haiti's Recent Exchange Rate Developments." IMF Working Paper WP/22/225.

IMF. (2022a). "Assessing Reserve Adequacy Database," www.imf.org/external/datamapper/datasets/ARA

Ismail, K, Perrelli, R, and Yang, J. (2020). "Optimism Bias in Growth Forecasts—The Role of Planned Policy Adjustments." IMF working paper WP/20/229

Itskhoki, O. (2020). "The Story of the Real Exchange Rate. NBER 28225

Kenen, P. B., and Schwartz, S. B. (1986). "An assessment of macroeconomic forecasts in the IMF's WEO." *Working Papers in International Economics*, Princeton University: International Finance Section.

Kentikelenis, A. E., Stubbs, T. H. and King L. P. (2016). "IMF conditionality and development policy space, 1985–2014." *Review of International Political Economy, 23,* 4, pp.543-582.

Luna, F. (2014). "IMF Forecasts in the Context of Program Countries." *Independent Evaluation Office BP/14/05*, Washington DC: IMF.

Mincer, J., and Zarnowitz V. (1969). "The Evaluation of Economic Forecasts." In J. Mincer (ed.) *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, pp. 3– 46. NY: NBER.

Mody, M. A. and Rebucci M. A. (2006). "Overview." in Mody, M. A. and Rebucci M.A. eds *IMF-supported programs: recent staff research*. Washington DC.: IMF.

Musso, A., and Phillips, S. (2002). "Comparing Projections and Outcomes of IMF-Supported Programs." *IMF Staff Papers*, 49, (1) pp. 22-48.

Nordhaus, W. (1987). "Forecasting efficiency: concepts and applications." *Review of Economics and Statistics*, 69, pp. 667–674.

Przeworski, A. and Vreeland, J. R. (2000). "The effect of IMF programs on economic growth." *Journal of Development Economics*, 62, (2), pp. 385-421.

Raftery, A. (1995). "Bayesian Model Selection in Social Research," Sociological Methodology, 25, pp. 111-163

Sinclair, T. M., Joutz, F., and Stekler, H. O. (2010). "Can the Fed predict the state of the economy?" *Economics Letters*, 108, (1), pp. 28-32.

Stubbs, T., Reinsberg, B., Kentikelenis, A., & King, L. (2020). "How to evaluate the effects of IMF conditionality." *The Review of International Organizations*, *15*(1), PP. 29-73.

Theil, H. (1961). "*Economic Forecasts and Policy*." Amsterdam: North-Holland.

Timmermann, A. (2007). "An Evaluation of the WEO Forecasts." *IMF Staff Papers*, *54*, (1), pp. 1–33.

U.S. Government Accountability Office (USGAO). (2003). "International financial crises: challenges remain in IMF's ability to anticipate, prevent, and resolve financial crises." GAO-03-734, Washington: Government Accountability Office.

Wijnholds, B and Kapteyn A. (2001). "Reserve Adequacy in Emerging Market Economies." IMF WP/01/143

World Bank. (2022a). "World Bank national accounts data, and OECD National Accounts data files." https://data.worldbank.org/indicator/NE.TRD.GNFS.ZS.

World Bank. (2022b). "Country Lending Group Data" databank.worldbank.org/data/download/site-content/OGHIST.xls

World Bank. (2022c) "World Bank International Debt Statistics." https://www.worldbank.org/en/programs/debt-statistics/ids

World Bank. (2022d). "World Development Indicators." https://databank.worldbank.org/source/world-development-indicators

Appendix A: Background Regressions for Figures

Table A1 Regressions for Figure 3a/b (Reserve Coverage)

Dependent Variable: Reserve Coverage	1996-1992	1997-1993	1998-1994	1999-1995	2000-1996	2001-1997	2002-1998	2003-1999	2004-2000	2005-2001	2006-2002	2007-2003	2008-2004	2009-2005	2010-2006	2011-2007	2012-2008	2013-2009	2014-2010	2015-2011	2016-2012	2017-2013	2018-2014	2019-2015
Forecast (β)	0.86	0.87	0.92	0.99	0.96	0.95	0.95	0.86	0.87	0.82	0.77	1	1.14	1.15	1.18	1.13	0.98	1.04	1.05	0.97	1.08	1.1	1.06	1.02
p-value (β=1)	0.13	0.11	0.34	0.86	0.68	0.54	0.60	0.13	0.18	0.15	0.10	0.99	0.37	0.36	0.22	0.39	0.87	0.76	0.74	0.85	0.47	0.47	0.64	0.87
Constant (α)	0.84***	0.88***	0.72**	0.40	0.48	0.61*	0.56	1.13**	1.08**	1.60***	1.80***	0.85	0.07	0.40	0.13	0.51	1.02**	0.88*	0.57	0.80	0.11	-0.04	0.17	0.43
p-value (α=0)	0.00	0.00	0.01	0.17	0.15	0.08	0.18	0.01	0.01	0.01	0.01	0.32	0.90	0.49	0.80	0.38	0.05	0.10	0.34	0.22	0.80	0.94	0.72	0.32
Observations	73	85	80	75	65	55	48	47	41	35	29	24	25	32	43	50	62	58	53	44	52	45	49	52
Adj. R^2	0.718	0.719	0.719	0.770	0.770	0.780	0.734	0.726	0.764	0.727	0.685	0.763	0.807	0.750	0.740	0.676	0.589	0.625	0.628	0.585	0.790	0.786	0.790	0.785
MZ p-value (α=0 & β=1)	0.00***	0.00***	0.00***	0.06*	0.10	0.06*	0.20	0.01**	0.01***	0.00***	0.00***	0.01***	0.11	0.01***	0.00***	0.00***	0.00***	0.00***	0.00***	0.01***	0.01***	0.04**	0.01**	0.00***
HP p-value (γ=0)	0.04**	0.01***	0.01***	0.04**	0.09*	0.04**	0.18	0.13	0.11	0.01**	0.07*	0.03**	0.05**	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.02**	0.01***	0.00***

Table A2 Regressions for Figure 3a/b (Reserve Levels)

Dependent Variable: Reserve Levels	1996-1992	1997-1993	1998-1994	1999-1995	2000-1996	2001-1997	2002-1998	2003-1999	2004-2000	2005-2001	2006-2002	2007-2003	2008-2004	2009-2005	2010-2006	2011-2007	2012-2008	2013-2009	2014-2010	2015-2011	2016-2012	2017-2013	2018-2014	2019-2015
Forecast (β)	0.99	1.05	1.04	0.79	0.78*	0.79	0.80	0.79	0.99	1.21*	1.23	1.38	1.32**	1.22*	1.14	1.08	1.06	1.04	1.01	0.94	0.99	0.98***	0.99	0.99
p-value (β=1)	0.83	0.42	0.43	0.16	0.08	0.12	0.14	0.13	0.93	0.07	0.12	000318*	0.02	0.05	00411*	0.22	0.31	0.41	0.80	0.16	0.78	0.00	0.33	0.44
Constant (α)	354.7***	278.3***	317.0**	1,361.0**	1,464.0**	1,557.1**	1,538.0*	1,754.0*	743.4*	7.57	41.28	-510.31	-367.79	193.97	368.45	630.18*	453.49*	460.25**	181.14	189.90	-74.64	-112.97	-3.95	323.88
p-value (α=0)	0.00	0.01	0.01	0.04	0.03	0.05	0.08	0.05	0.08	0.99	0.96	0.32	0.62	0.70	0.14	0.05	0.05	0.01	0.37	0.34	0.67	0.73	0.99	0.33
Observations	73	85	80	75	65	55	48	47	41	35	29	24	25	32	43	50	62	58	53	44	52	45	49	52
Adj. R^2	0.900	0.912	0.935	0.856	0.869	0.863	0.872	0.862	0.936	0.956	0.948	0.961	0.925	0.937	0.960	0.956	0.964	0.985	0.965	0.961	0.952	0.994	0.992	0.995
MZ p-value (α=0 & β=1)	0.00***	0.00***	0.01***	0.05*	0.09*	0.12	0.21	0.12	0.03**	0.00***	0.00***	0.00***	0.01***	0.01***	0.00***	0.01***	0.01***	0.00***	0.54	0.30	0.83	0.00***	0.62	0.47
HP p-value (γ=0)	0.10	0.02**	0.02**	0.85	0.67	0.74	0.55	0.72	0.09*	0.00***	0.01***	0.01***	0.01**	0.01**	0.00***	0.01***	0.01***	0.00***	0.40	0.47	0.60	0.28	0.64	0.59

Table A3 Regressions for Figure 5 (Short-Term External Debt)

Dependent Variable: Short Term Debt	1996-1992	1997-1993	1998-1994	1999-1995	2000-1996	2001-1997	2002-1998	2003-1999	2004-2000	2005-2001	2006-2002	2007-2003	2008-2004	2009-2005	2010-2006	2011-2007	2012-2008	2013-2009	2014-2010	2015-2011	2016-2012	2017-2013	2018-2014	2019-2015
Forecast (β)	2.5***	2.49***	2.51***	1.17	0.85***	1.00	1.08	0.93	0.96	1.32	1.33	1.35	1.51	1.44	1.08	1.00	0.98	0.72**	0.73*	0.63***	0.35***	1.39	1.00	0.98
p-value (β=1)	0.00	0.00	0.00	0.69	0.00	0.98	0.70	0.65	0.83	0.33	0.38	0.43	0.17	0.21	0.75	0.98	0.93	0.02	0.09	0.00	0.00	0.55	0.99	0.94
Constant (α)	-0.80	-0.78	-1.00	1.74	1.20***	0.90**	2.09*	2.00**	2.04**	0.64	0.69	0.17	0.15	-1.10	0.48	0.52	0.78*	1.12***	1.43***	1.98***	3.05***	0.55	3.09*	3.31**
p-value (α=0)	0.14	0.20	0.35	0.16	0.00	0.02	0.08	0.03	0.01	0.57	0.58	0.89	0.88	0.35	0.37	0.22	0.09	0.01	0.00	0.01	0.00	0.75	0.06	0.04
Observations	14	13	8	10	8	9	12	20	23	26	22	19	16	22	29	33	40	39	32	25	29	25	28	29
Adj. R^2	0.934	0.931	0.927	0.632	0.994	0.952	0.736	0.734	0.720	0.769	0.749	0.778	0.824	0.795	0.779	0.768	0.755	0.766	0.727	0.609	0.271	0.515	0.513	0.551
MZ p-value (α=0 & β=1)	0.00***	0.00***	0.00***	0.32	0.00***	0.04**	0.20	0.09*	0.03**	0.05*	0.10	0.15	0.07*	0.44	0.17	0.21	0.12	0.01***	0.01**	0.00***	0.00***	0.39	0.15	0.12
HP p-value (γ=0)	0.15	0.15	0.24	0.29	0.36	0.25	0.19	0.31	0.15	0.09*	0.13	0.24	0.13	0.27	0.34	0.57	0.35	0.88	0.39	0.45	0.67	0.38	0.44	0.43

Robust p values, *** p<0.01, ** p<0.05, * p<0.1

Table A4: Mincer Zarnowitz Regressions for Figures 5a and 5b
Forecast Accuracy By Month

	Reserve Coverage											
	Jan	**Feb**	**Mar**	**Apr**	**May**	**June**	**July**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
MONA Forecast, β	0.91	1.11	0.76**	0.77*	0.77*	1.06	1.16	0.89	0.95	0.99	0.87	1.21**
p-value ($\beta = 1$)	0.57	0.55	0.02	0.06	0.07	0.27	0.24	0.56	0.53	0.96	0.61	0.01
Constant, (α)	1.01*	0.41	1.20***	1.48***	1.47**	0.14	0.38	0.60	0.46	0.75*	0.81	-0.22
p-value ($\alpha = 0$)	0.08	0.66	0.00	0.01	0.02	0.68	0.50	0.31	0.19	0.06	0.24	0.64
Observations	30	16	32	30	25	22	39	16	19	11	19	28
Adj. R-sq	0.70	0.77	0.69	0.41	0.59	0.91	0.72	0.66	0.92	0.74	0.57	0.88
MZ p-value (α=0, β=1)	0.02**	0.08*	0.01**	0.01**	0.07*	0.07*	0.00***	0.47	0.29	0.00***	0.13	0.03**
HP p-value (γ=0)	0.08*	0.03**	0.81	0.01***	0.08*	0.04**	0.00***	0.68	0.32	0.00***	0.29	0.02**
	Reserves											
	Jan	**Feb**	**Mar**	**Apr**	**May**	**June**	**July**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
MONA Forecast, β	0.88	1.12***	0.85**	0.89	1.11	1.17***	1.06	1.05	1.04	1.09***	0.99	1.04
p-value ($\beta = 1$)	0.30	0.01	0.01	0.60	0.29	0.00	0.12	0.54	0.15	0.00	0.20	0.59
Constant, (α)	1,046.50	197.82	354.12	803.58*	324.55	-415.29*	725.13**	-26.96	183.89	62.14*	1,256.29	172.57
p-value ($\alpha = 0$)	0.13	0.57	0.25	0.09	0.25	0.07	0.03	0.95	0.48	0.05	0.19	0.52
Observations	30	16	32	30	25	22	39	16	19	11	19	28
Adj. R-sq	0.93	0.99	0.97	0.75	0.96	0.99	0.96	0.93	0.99	1.00	1.00	0.95
MZ p-value (α=0, β=1)	0.30	0.01***	0.03**	0.05**	0.14	0.00***	0.02**	0.78	0.24	0.00***	0.26	0.52
HP p-value (γ=0)	0.81	0.02**	0.10*	0.48	0.13	0.33	0.00***	0.51	0.12	0.00***	0.31	0.30

Robust p values. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix B: Data Sources

Variable	Data Source	Description / Database Codenames
Current-Year Forecasts	IMF MONA (IMF 2021a). For MONA audit and error correction, see Appendix B	**Total gross reserves in per USD** Pre 2002 MONA data o **IMF MONA**: FAFA o If FAFA is not available, collected from EBS Documents Post-2002 MONA data: o Collected from EBS Documents **Imports of goods and services in USD** Pre 2002 MONA data o **IMF MONA**: BMT+ BMS_O o If BMS_O is not available, **IMF MONA**: BMG Post-2002 MONA data: o **IMF MONA**: BMGS o If BMGS is not available, **IMF MONA**: BMG +BMS o IF BMGS, BMG, BMS are not available, **IMF MONA**: NM/ENDA **Short-term external debt in USD** Pre 2002 MONA data: o **IMF MONA**: Collected from EBS Documents Post-2002 MONA dataset: o **IMF MONA:** D_S, o If D_S is not available, **IMFMONA**: D - D_L o If D_S, D - D_L are not available, collected from EBS Documents
Actual Final Data	IMF IFS, (IMF 2021c); IMF BOP, (IMF 2021d); IMF ARA, (IMF 2022a); WB WDI, (WB 2021a).	**Total gross reserves in USD** **WB WDI**: FI.RES.TOTL.CD o If FI.RES.TOTL.CD is not available, **IMF ARA**: Gross Reserve in USD **Imports of goods and services in USD** o **IMF IFS**: BMGS_BP6_USD, o If BMGS_BP6_USD is not available, **IMF BOP**: BM.GSR.GNFS.CD **Short-term external debt in USD** o **WB WDI**: DT.DOD.DSTC.CD o If DT.DOD.DSTC.CD is not available, **IMF ARA**: Short-Term external debt in USD
Elections Data	Beck et al. (2001), IFES (2020)	Election dummy for head of state, government, legislative elections. Program received a "1" if election occurred up to 1 year prior to program start. Details at electionguide.org. Pre 1998, Beck et al data.
Conflicts Data	Harbom et al. (2009)	Conflict dummy covers intra-state & inter-state conflicts. Program received a "1" if country experienced a conflict up to one year prior to program start date.
Disasters Data	EM-DAT (2020)	Disaster dummy covers natural disasters. Program received a "1" if a disaster occurred up to 1 year prior to the program start date.
Condi-tionality	IMF MONA, (IMF 2021a).	Dummy defined by MONA's Glossary (IMF 2021a) for quantitative performance targets: Domestic Credit Ceiling, Gov't/Public Sector Credit Ceilings, BOP Reserve Tests, Debt Ceilings (short, medium and long term), Arrears Ceilings (domestic and external), Fiscal Deficit Ceilings.
CRISES	endogenous	1997 Asian Crisis: dummy for programs approved 5/15/1997-3/25/1999 (the last Indonesia program). 2008 Crisis: dummy for programs approved 9/15/2007-12/31/2008.
CRISES	exogenous	1997 Crisis: dummy for programs approved one year before 5/15/1997; 2008 Crisis: dummy programs approved one year before 9/15/2007 (and duration overlapped with the advent of the crisis)

Appendix C: Auditing the MONA Database

The MONA database presents challenges as it contains a wide range of errors. Unlike the WEO database, MONA does not include release dates, hence it is unclear if/when revisions/updates to the database take place. To prevent data errors from deriving our results, we audited MONA and corrected the following 9 different kinds of errors that fall into three major categories:

Data Entry Errors
C.1. Data Entered with Wrong Signs
C.2. Temporal Errors: Correct Data Entered for the Wrong Program Year
C.3. Zeros Identify Missing Values
C.4. Typos and Spelling Mistakes
C.5. Wrong Line Items Entered
Inconsistencies
C.6. Currency Unit Inconsistencies
C.7. Unit Magnitude Inconsistencies
Corrected Data from IMF Archives (Executive Board Documents)
C.8. Missing Data Corrected
C.9. Outliers Corrected

C.1. Data Entered with Incorrect Signs
We corrected 1500 observations that had been entered with incorrect signs. Most errors affect trade data. Imports and Exports are supposed to be entered with positive signs, for example, but many imports are accidentally entered with negative signs.

Table C.1 Data with Incorrect Sign

Count of Total Corrections	t-3	t-2	t-1	t	t+1	t+2	t+3	t+4
1500	160	185	153	165	194	215	217	211

C.2: Temporal Errors
MONA reports data from *t-3* to *t+4*, where "*t*" is the program year. For example, if the program year is 1997, then MONA reports data from 1994 to 2001 and "*t*" should data for 1997. Sometimes, data entry confused the program year and generated temporal errors associating the correct data with the wrong program year (e.g. 1997 data is entered in the *t+1* spot). Seven programs suffered this error.

Table C.2 Temporal Errors

Count	Prog.	Country	Year	Mnemonic	Review	Correction
1	7	ESTONIA	1993	All	Last	Corrected data using IMF archives
2	15	EL SALVADOR	1993	All	Last	Corrected data using IMF archives
3	552	DOMINICA	2005	All	All	Corrected data using IMF archives
4	579	GABON	2007	All	R1-Last	Corrected data using IMF archives

C.3: Zeros Identify Missing Values
MONA does not possess a consistent indicator for missing values. At times missing values are presented as "NA", ".", "0", or "NULL". There are over 2500 entries in MONA with values of exact zero for variables such as levels of imports, GDP, exchange rates, inflation. This is suspicious data that we surmise represents missing observations - although STdebt data of zero may well be accurate. Whenever a zero is observed in MONA for debt levels, we cross referenced the data with the source data in IMF archival loan documents. If the zero could not be verified, and if the actual final outcome data was not zero, we assumed the zero indicates a missing value by erring on the side of caution.

Table C.3 Zeros Assumed to Indicate Missing Data Since Zero Could Not Be Verified In Loan Documents (Although Final Outcome Data Is Not Zero)

Count	Prog.	Country	Year	Mnemonic	Review
1	506	BOSNIA HERZEGOVINA	2002	D_S	R0
2	546	UKRAINE	2004	D_S	R0
3	564	IRAQ	2005	D_S	R0
4	568	PARAGUAY	2006	D_S	R0
5	579	GABON	2007	D_S	R0
6	588	IRAQ	2007	D_S	R0
7	601	SEYCHELLES	2008	D_S	R0
8	603	ICELAND	2008	D_S	R0
9	611	ARMENIA	2009	D_S	R0
10	626	SEYCHELLES	2010	D_S	R0
11	627	MALAWI	2009	D_S	R0
12	628	KYRGYZ REP	2008	D_S	R0
13	629	ETHIOPIA	2009	D_S	R0
14	632	MALAWI	2010	D_S	R0
15	633	IRAQ	2010	D_S	R0
16	640	SOLOMON ISLANDS	2010	D_S	R0
17	641	LESOTHO	2010	D_S	R0
18	643	SIERRA LEONE	2010	D_S	R0
19	649	ARMENIA	2010	D_S	R0
20	650	KOSOVO, REP OF	2010	D_S	R0
21	651	HAITI	2010	D_S	R0
22	653	YEMEN	2010	D_S	R0
23	661	KENYA	2011	D_S	R0
24	671	KYRGYZ REP	2011	D_S	R0
25	675	AFGHANISTAN	2011	D_S	R0
26	676	SOLOMON ISLANDS	2011	D_S	R0
27	684	KOSOVO, REP OF	2012	D_S	R0
28	685	GAMBIA, THE	2012	D_S	R0
29	686	CENTRAL AFRICAN REP	2012	D_S	R0
30	687	TANZANIA	2012	D_S	R0
31	688	SAO TOME / PRINCIPE	2012	D_S	R0
32	689	MALAWI	2012	D_S	R0
33	693	LIBERIA	2012	D_S	R0
34	697	JAMAICA	2013	D_S	R0
35	709	ALBANIA	2014	D_S	R0
36	711	UKRAINE	2014	D_S	R0
37	712	SEYCHELLES	2014	D_S	R0
38	717	CHAD	2014	D_S	R0
39	722	KENYA	2015	D_S	R0
40	725	GHANA	2015	D_S	R0
41	726	KYRGYZ REP	2015	D_S	R0
42	732	KOSOVO, REP OF	2015	D_S	R0
43	733	MOZAMBIQUE	2016	D_S	R0
44	734	734KENYA	2016	D_S	R0
45	739	RWANDA	2016	D_S	R0
46	741	IRAQ	2016	D_S	R0
47	742	MADAGASCAR	2016	D_S	R0
48	746	CENTRAL AFRICAN REP	2016	D_S	R0
49	748	JAMAICA	2016	D_S	R0
50	757	MONGOLIA	2017	D_S	R0
51	758	SIERRA LEONE	2017	D_S	R0
52	759	GABON	2017	D_S	R0
53	765	GUINEA	2017	D_S	R0
54	772	BARBADOS	2018	D_S	R0
55	773	SIERRA LEONE	2018	D_S	R0
56	784	CONGO, REP OF	2019	D_S	R0
57	786	SAO TOME / PRINCIPE	2019	D_S	R0
58	788	EQUATORIAL GUINEA	2019	D_S	R0
59	789	LIBERIA	2019	D_S	R0
60	790	CENTRAL AFRICAN REP	2019	D_S	R0
61	791	ETHIOPIA	2019	D_S	R0

C.4: Typos and Spelling Mistakes

We adjusted 18 observations when (i) a series is misspelled, (ii) one decimal is incorrect, (iii) one additional integer is added in the wrong place, (iv) one integer is missing, (v) the wrong country is being identified as the program country, (vi) the wrong year is identified as the program year, or (vii) when the variable contained typos. These typos and spelling errors were corrected based on the original IMF *Executive Board Special* (EBS) loan documents.

Table C.4 Typos and Spelling Mistakes

Count	Prog.	Country	Year	Mnemonic	Review	Correction
1	15	EL SALVADOR	1993	programyear	R2	programyear corrected
2	18	LATVIA	1993	programyear	R1-Last	programyear corrected
3	117	ALBANIA	1994	countryname	All	wrong countryname corrected
4	117	ALBANIA	1994	countryncode	All	wrong countryname corrected
5	132	SIERRA LEONE	1995	programyear	R0	programyear corrected
6	143	PAKISTAN	1996	programyear	R0-R1	programyear corrected
7	205	VIETNAM	1996	boarddocno	R1	board document typo corrected
8	207	ETHIOPIA	1997	reviewtype	All	reviewtype labels corrected
9	230	BURKINA FASO	1996	programyear	R0	Corrected data using IMF archives
10	274	UKRAINE	1998	programyear	R5-R6	programyear corrected
11	527	NICARAGUA	2002	programyear	R10	programyear corrected
12	560	BENIN	2005	boarddocno	R0	board document typo corrected
13	628	KYRGYZ REP	2008	reviewtype	All	reviewtype labels corrected
14	681	NIGER	2012	programyear	R8	programyear corrected
15	724	UKRAINE	2015	reviewtype	All	reviewtype labels corrected
16	734	KENYA	2016	reviewtype	All	reviewtype labels corrected
17	764	MAURITANIA	2017	programyear	R0-R4	programyear corrected
18	All	All	All	initialenddate	All	spelling error corrected

C.5: Wrong Line Item Entered

At times, data entry inadvertently fell into the wrong line and entered the wrong line item. For example, instead of entering reserve data, data entry entered GDP data from one line above/below the reserve data.

Table C.5 Wrong Line Item Entered

Count	Prog.	Country	Year	Mnemonic	Review	Correction
1	518	DOMINICA	2002	G_RES	R0	Corrected data using IMF archives
2	579	GABON	2007	G_RES	R0	Corrected data using IMF archives
3	600	HUNGARY	2008	G_RES	R0	Corrected data using IMF archives
4	603	ICELAND	2008	G_RES	R0	Corrected data using IMF archives

C.6: Inconsistent Currency Units
We corrected 27 instances when the currency in MONA was incorrectly identified, as verified by IMF archival loan documents.

Table C.6 Incorrect Currency Unit

Count	Prog.	Country	Year	Mnemonic	Review	Correction
1	15	EL SALVADOR	1993	D_S	R0	Corrected to USD using IMF archives
2	78	SLOVAK REP	1994	D_S	R0-R1	Corrected to USD using IMF archives
3	91	EL SALVADOR	1995	D_S	R0	Corrected to USD using IMF archives
4	112	CHAD	1994	RES_G	All	Corrected to USD using IMF archives
5	118	CONGO, REP	1994	RES_G	All	Corrected to USD using IMF archives
6	136	HAITI	1995	D_S	R0	Corrected to USD using IMF archives
7	159	AZERBAIJAN	1996	RES_G	All	Corrected to USD using IMF archives
8	162	GABON	1995	RES_G	All	Corrected to USD using IMF archives
9	248	CAPE VERDE	1998	RES_G	All	Corrected to USD using IMF archives
10	375	PAKISTAN	2000	D_S	R0	Corrected to USD using IMF archives
11	517	CROATIA	2003	D_S	All	Corrected to USD using IMF archives
12	523	GUATEMALA	2003	D_S	All	Corrected to USD using IMF archives
13	539	DOMINICAN REP	2003	D_S	R0	Corrected to USD using IMF archives
14	548	CROATIA	2004	D_S	R0-R1	Corrected to USD using IMF archives
15	549	BULGARIA	2004	RES_G	All	Corrected to USD using IMF archives
16	552	DOMINICAN REP	2005	D_S	R0	Corrected to USD using IMF archives
17	579	GABON	2007	RES_G	All	Corrected to USD using IMF archives
18	606	SERBIA,REP	2009	RES_G	All	Corrected to USD using IMF archives
19	613	MONGOLIA	2009	D_S	R0	Corrected to USD using IMF archives
20	617	ROMANIA	2009	RES_G	All	Corrected to USD using IMF archives
21	662	ROMANIA	2011	RES_G	All	Corrected to USD using IMF archives
22	673	SERBIA,REP	2011	RES_G	All	Corrected to USD using IMF archives
23	692	BOSNIA HERZEGO.	2012	RES_G	All	Corrected to USD using IMF archives
24	704	ROMANIA	2013	RES_G	All	Corrected to USD using IMF archives
25	723	SERBIA,REP	2015	RES_G	All	Corrected to USD using IMF archives
26	742	MADAGASCAR	2016	RES_G	All	Corrected to USD using IMF archives
27	760	CAMEROON	2017	D_S	R0	Corrected to USD using IMF archives

C.7: Unit Magnitude Inconsistencies
For example, when MONA indicates data in millions when data is actually in billion or thousands.

Table C.7 Base Year Errors

Count	Prog.	Country	Year	Mnemonic	Review	Correction
1	16	KYRGYZ REP	1993	All	All	Unresolved, dropped
2	108	KAZAKSTAN	1994	All	All	Unresolved, dropped
3	160	RUSSIAN FED	1995	FAFA	R0	Corrected based on EBS
4	164	RUSSIAN FED.	1996	FAFA	R0	Corrected based on EBS
5	222	MEXICO	1995	FAFA	R0	Corrected based on EBS
6	242	BOSNIA HERZEGOVINA	1998	FAFA	R0	Unresolved, dropped
7	248	CAPE VERDE	1998	FAFA	R0	Unresolved, dropped
8	255	THAILAND	1997	FAFA	R0	Corrected based on EBS
9	256	INDONESIA	1998	FAFA	R0	Corrected based on EBS
10	275	INDONESIA	1999	FAFA	R0	Corrected based on EBS
11	302	RUSSIAN FED.	1999	FAFA	R0	Corrected based on EBS
12	308	BRAZIL	1999	FAFA	R0	Corrected based on EBS
13	328	ARGENTINA	2000	FAFA	R0	Corrected based on EBS
14	337	INDONESIA	2000	FAFA	R0	Corrected based on EBS
15	401	BRAZIL	2001	FAFA	R0	Corrected based on EBS
16	644	ANTIGUA & BARBUDA	2010	G_RES	R0	Unresolved, dropped

C.8: Missing Data

Missing data encountered in the MONA database was filled in using the IMF Archives' *Executive Board Special* (EBS) loan documents when available. We filled in 49 observations listed below.

Table C.8: Missing MONA Data Filled Using IMF Archives

Count	Prog.	Country	Year	Mnemonic	Review	Correction
1	4	COSTA RICA	1993	D_S_t	R0	Entered data using IMF archives
2	9	EGYPT	1993	D_S_t	R0	Entered data using IMF archives
3	14	JAMAICA	1992	D_S_t	R0	Entered data using IMF archives
4	15	EL SALVADOR	1993	D_S_t	R0	Entered data using IMF archives
5	23	PAKISTAN	1993	D_S_t	R0	Entered data using IMF archives
6	24	PERU	1993	D_S_t	R0	Entered data using IMF archives
7	75	TURKEY	1994	D_S_t	R0	Entered data using IMF archives
8	78	SLOVAK REPUBLIC	1994	D_S_t	R0	Entered data using IMF archives
9	83	PAKISTAN	1993	D_S_t	R0	Entered data using IMF archives
10	91	EL SALVADOR	1995	D_S_t	R0	Entered data using IMF archives
11	112	CHAD	1994	G_Res	R0	Entered data using IMF archives
12	122	ZIMBABWE	1992	D_S_t	R0	Entered data using IMF archives
13	136	HAITI	1995	D_S_t	R0	Entered data using IMF archives
14	146	HUNGARY	1996	D_S_t	R0	Entered data using IMF archives
15	195	PERU	1996	D_S_t	R0	Entered data using IMF archives
16	217	VENEZUELA	1996	D_S_t	R0	Entered data using IMF archives
17	220	EL SALVADOR	1997	D_S_t	R0	Entered data using IMF archives
18	222	MEXICO	1995	D_S_t	R0	Entered data using IMF archives
19	242	BOSNIA HERZEGO.	1998	G_Res	R0	Entered data using IMF archives
20	248	CAPE VERDE	1998	G_Res	R0	Entered data using IMF archives
21	276	KOREA	1997	D_S_t	R0	Entered data using IMF archives
22	293	EL SALVADOR	1998	D_S_t	R0	Entered data using IMF archives
23	300	MEXICO	1999	D_S_t	R0	Entered data using IMF archives
24	304	JORDAN	1999	D_S_t	R0	Entered data using IMF archives
25	339	URUGUAY	2000	D_S_t	R0	Entered data using IMF archives
26	356	SRI LANKA	2001	D_S_t	R0	Entered data using IMF archives
27	375	PAKISTAN	2000	D_S_t	R0	Entered data using IMF archives
28	400	LATVIA	2001	D_S_t	R0	Entered data using IMF archives
29	401	BRAZIL	2001	D_S_t	R0	Entered data using IMF archives
30	404	LITHUANIA	2001	D_S_t	R0	Entered data using IMF archives
31	418	TURKEY	2002	D_S_t	R0	Entered data using IMF archives
32	512	BOLIVIA	2003	D_S_t	R0	Entered data using IMF archives
33	528	PARAGUAY	2003	D_S_t	R0	Entered data using IMF archives
34	531	ROMANIA	2004	D_S_t	R0	Entered data using IMF archives
35	540	GABON	2004	D_S_t	R0	Entered data using IMF archives
36	546	UKRAINE	2004	D_S_t	R0	Entered data using IMF archives
37	562	N. MACEDONIA REP	2005	D_S_t	R0	Entered data using IMF archives
38	600	HUNGARY	2008	G_Res	R0	Entered data using IMF archives
39	616	GUATEMALA	2009	G_Res	R0	Entered data using IMF archives
40	623	ANGOLA	2010	D_S_t	R0	Entered data using IMF archives
41	636	GRENADA	2010	D_S_t	R0	Entered data using IMF archives
42	672	ST. KITTS & NEVIS	2011	D_S_t	R0	Entered data using IMF archives
43	678	BURUNDI	2012	D_S_t	R0	Entered data using IMF archives
44	679	GUINEA	2012	D_S_t	R0	Entered data using IMF archives
45	705	SIERRA LEONE	2013	D_S_t	R0	Entered data using IMF archives
46	715	MOROCCO	2014	D_S_t	R0	Entered data using IMF archives
47	724	UKRAINE	2015	D_S_t	R0	Entered data using IMF archives
48	745	AFGHANISTAN, REP	2016	D_S_t	R0	Entered data using IMF archives
49	791	ETHIOPIA	2019	G_Res	R0	Entered data using IMF archives

C.9: Outliers Verification / Correction

We audited observations that fell three or more standard deviations from the mean. Then we recalculated the distribution and we conducted a second rounds of outlier checks. Outliers were checked using the original IMF *Executive Board Special* (EBS) loan documents. We corrected 18 observations using the EBS, as listed below.

Table C.9 Outliers Corrected

Count	Prog.	Country	Year	Mnemonic	Review	Correction
1	19	LITHUANIA	1993	FAFA	R0	Entered data using IMF archives
2	23	PAKISTAN	1993	FAFA	R0	Entered data using IMF archives
3	24	PERU	1993	FAFA	R0	Entered data using IMF archives
4	34	CTR AFRICAN REP	1994	FAFA	R0	Entered data using IMF archives
5	65	MALAWI	1994	FAFA	R0	Entered data using IMF archives
6	143	PAKISTAN	1996	FAFA	R0	Entered data using IMF archives
7	187	UZBEKISTAN	1996	FAFA	R0	Entered data using IMF archives
8	294	BULGARIA	1998	FAFA	R0	Entered data using IMF archives
9	374	GABON	2000	FAFA	R0	Entered data using IMF archives
10	398	BULGARIA	2002	FAFA	R0	Entered data using IMF archives
11	505	JORDAN	2002	D_S	R0	Entered data using IMF archives
12	552	DOMINICAN REP.	2005	D_S	R0	Entered data using IMF archives
13	564	IRAQ	2005	G_RES	R0	Entered data using IMF archives
14	576	PERU	2007	D_S	R0	Entered data using IMF archives
15	604	PAKISTAN	2008	G_RES	R0	Entered data using IMF archives
16	606	SERBIA, REP OF	2009	D_S	R0	Entered data using IMF archives
17	608	BELARUS	2009	G_RES	R0	Entered data using IMF archives
18	620	SRI LANKA	2009	D_S	R0	Entered data using IMF archives